MARCELA MENDES SALAZAR

ANOTAÇÃO E CARACTERIZAÇÃO PRELIMINAR DE GENES DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DE *Eucalyptus*

Dissertação apresentada ao Programa de Pós-graduação em Biologia Molecular e Celular, da Universidade Federal de Goiás, como requisito à obtenção do título de Mestre em Biologia Molecular e Celular.

Goiânia, GO - Brasil

2006

MARCELA MENDES SALAZAR

ANOTAÇÃO E CARACTERIZAÇÃO PRELIMINAR DE GENES DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DE *Eucalyptus*

Orientador:

Prof. Dr. Alexandre Siqueira Guedes Coelho

Co-Orientadora:

Profa. Dra. Rosana Pereira Vianello Brondani

Goiânia, GO - Brasil

2006

Aos meus pais, Sidnei Salazar e Regina Salazar, e à minha avó Olga Leão, dedico.

Aos meus avós, Mário Salazar (*in memorian*) e Thereza Salazar, ofereço.

AGRADECIMENTOS

Agradeço, em primeiro lugar, à minha mãe pelo grandioso esforço para que eu chegasse até aqui. E ao meu pai, por realmente acreditar que eu chegaria. Pessoas que amo tanto e que são o meu orgulho e referencial de caráter.

Aos meus avós, Olga Leão, Thereza e Mário Salazar (*in memorian*) por todo carinho e dedicação.

Ao admirável profissional Dr. Alexandre Siqueira pela orientação e paciência nesses anos de convivência.

A Dra. Rosana Brondani da Embrapa Arroz e Feijão pela co-orientação e pelas suas valiosas contribuições para o êxito desse trabalho.

A todos os professores do programa de Mestrado em Biologia Molecular e Celular da UFG pelas grandes contribuições ao meu conhecimento em Biologia. E à Capes, pela bolsa.

Aos amigos do laboratório de Genética e Genômica de Plantas da UFG: Ana Clara, Angel, Carol, Daniela, Dany Miranda, Elaine, Ludmila, Márcia, Marlei, Rafael, Raylla, por tornarem o trabalho tão prazeroso. Gostaria de agradecer em especial à Elaine por compartilhar e colaborar com esse trabalho, ao Angel e Lud por estarem sempre dispostos a esclarecerem minhas dúvidas.

Às colegas do laboratório de Biotecnologia da Embrapa Arroz e Feijão em especial à Tereza e Ana Rosa pelas preciosas contribuições.

À minha tão essencial família, principalmente aos meus irmãos, Rodrigo e Georgios, e aos primos-amigos Ariane e Rodollfo por tantos anos de companheirismo e momentos de descontração.

Aos amigos do Projeto Hubuntu, iniciativa que foi fundamental para consolidar alguns conceitos que tenho sobre a vida.

Aos amigos e médicos do Hospital da Visão pela compreensão e incentivo.

A todas as minhas amigas e amigos que são tão especiais e fundamentais na minha vida e de uma maneira muito especial à Karine por aturar tantas reclamações e por saber contorná-las de um jeito muito peculiar.

Agradeço, enfim, a Deus por permitir que todos esses “grandes ombros” estivessem sempre por perto para que eu pudesse me apoiar!

“Eu queria não saber que o saber é infinito,
não sofreria tanto pela minha ignorância.”

Sídnei Salazar

“O êxtase de se realizar alguma tarefa não
consiste apenas em se ver a obra pronta,
mas principalmente no que se aprendeu e
evoluiu no processo de sua construção.”

Sídnei Salazar

SUMÁRIO

Lista de Figuras VI
Lista de Tabelas X
Lista de Siglas XI
Resumo XII
Abstract XIV
1 INTRODUÇÃO 01
2 OBJETIVOS 04
3 REVISÃO DE LITERATURA 05
3.1 O GÊNERO *Eucalyptus* 05
3.1.1 Caracterização de algumas espécies do gênero *Eucalyptus* 07
3.2 A PAREDE CELULAR 09
3.2.1 Estrutura da parede celular 09
3.2.2 Estrutura e biossíntese de celulose 11
3.2.3 Genes *CesA* 12
3.2.4 Genes *AtCesA7* 14
3.2.5 Polimorfismos de um único nucleotídeo 15
4 MATERIAL E MÉTODOS 18
4.1 MINERAÇÃO DOS DADOS 18
4.2 DESENHO DE PRIMERS 21
4.3 TRIAGEM DE UMA BIBLIOTECA DE BACs EM BUSCA DE CLONES CONTENDO GENES HOMÓLOGOS A *AtCesA7* 22
4.4 ISOLAMENTO DOS BACs COTENDO O GENE DE CELULOSE SINTASE DE *Eucalyptus* HOMÓLOGO A *AtCesA7* 24
4.5 SEQÜENCIAMENTO DO GENE DE CELULOSE SINTASE DE *Eucalyptus* HOMÓLOGO A *AtCesA7* 24
4.6 ANÁLISE DAS SEQÜÊNCIAS 26
4.7 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DO GENE DE CELULOSE SINTASE DE *Eucalyptus* ORTÓLOGO A *AtCesA7* 27
4.8 DETECÇÃO DE POLIMORFISMO NO GENE *CesA7.2* DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO *Eucalyptus* 28
5 RESULTADOS E DISCUSSÃO 30
5.1 MINERAÇÃO DOS DADOS 30

5.2 MINERAÇÃO E DESENHO DE *PRIMERS* PARA GENES ORTÓLOGOS A *AtcesA4* 37
5.3 MINERAÇÃO E DESENHO DE *PRIMERS* PARA GENES ORTÓLOGOS A *AtcesA7* 39
5.4 MINERAÇÃO E DESENHO DE *PRIMERS* PARA GENES ORTÓLOGOS A *AtcesA8* 49
5.5 TRIAGEM DA BIBLIOTECA DE BACS PARA CLONES CONTENDO FRAGMENTOS DE CesA7.2 52
5.6 TRIAGEM DA BIBLIOTECA DE BACS COM OS PRIMERS CesA72.3 E CesA72.4 53
5.7 TRIAGEM DA BIBLIOTECA DE BACS COM OS PRIMERS EUC1.3 e EUC1.4 54
5.8 TRIAGEM DA BIBLIOTECA DE BACS COM OS PRIMERS EUC2.3 e EUC2.4 57
5.9 CONFIRMAÇÃO DA AMPLIFICAÇÃO DE FRAGMENTOS DIRETAMENTE NOS CLONES COM OS PRIMERS CesA72.5 E CesA72.6 59
5.10 ISOLAMENTO DO BACs POR LISE ALCALINA E AMPLIFICAÇÃO DOS FRAGMENTOS 61
5.11 SEQÜÊNCIAMENTO PARCIAL DO GENE CesA7.2 DE Eucalyptus grandis 62
5.12 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DO GENE CesA7.2 DE CELULOSE SINTASE DE Eucalyptus 65
5.13 DETECÇÃO DE POLIMORFISMO NO GENE CesA7.2 DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO Eucalyptus 67
6 CONSIDERAÇÕES FINAIS 75
7 REFERÊNCIAS BIBLIOGRÁFICAS 77
8 ANEXOS 86

LISTA DE FIGURAS

Figura 01. Estrutura da parede celular de plantas, evidenciando as diversas camadas que formam a parede celular secundária. Adaptado de Lodish et al. (2000). 10

Figura 02. Estrutura da interação das cadeias de celulose. Adaptado de Delmer & Amor (2000) 12

Figura 03. Estrutura em roseta do complexo celulose sintase. Os retângulos brancos representam cadeias de celulose recém-formadas. Adaptado de Brown Jr. e Saxena (2000) 13

Figura 04. Secção transversal do sistema vascular de *A. thaliana* evidenciando elementos do xilema no tipo selvagem (A) e no mutante *irx3* (B) (ph-floema). Adaptada de Brown et al. (2005) 15

Figura 05. Estrutura em hierarquia de placas, *pools* e *superpools* da biblioteca de *Eucalyptus grandis.* 22

Figura 06. Dendrograma de referência construído utilizando o algorítmo de Neighbor-Joining representando as relações entre genes de celulose sintase de *Arabidopsis thaliana* (Tabela 1). 31

Figura 07. Dendrogramas representativos das relações entre ESTs de *Eucalyptus* e seqüências de celulose sintase de outras espécies, utilizadas como referência (Tabelas 1 e 2), dendrograma resumido (A); grupo 1 (B); grupo 2 (C); grupo 3 (D); grupo 4 (E); grupo 5 (F) e grupo 6 (G) 32

Figura 08. Posição relativa do íntron no alinhamento dos ESTs ortólogos ao gene *AtCesA4*, inferida pela utilização do programa EST2GENOME. 37

Figura 09. Representação esquemática dos resultados obtidos pelo programa EST2GENOME para a localização das regiões correspondentes a éxons e íntrons no alinhamento dos ESTs ortólogos ao gene *AtCesA4.* A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA4*, as barras azuis representam o *contig* de ESTs de *Eucalyptus* ortólogas a esse gene. As setas indicam a localização e o sentido dos *primers* desenhados 39

Figura 10. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs ortólogos ao gene *AtCesA7*, grupo denominado de *CesA7.1*, inferida pela utilização do programa EST2GENOME 40

Figura 11. Representação esquemática do alinhamento realizado com o programa EST2GENOME para localização dos íntrons e éxons. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o primeiro grupo de ESTs de *Eucalyptus* ortólogos a esse gene (grupo *CesA7.1*). As setas indicam a localização e o sentido dos *primers* desenhados 41

Figura 12. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs ortólogos ao gene *AtCesA7*, grupo denominado de *CesA7.2*, inferida pela utilização do programa EST2GENOME 42

Figura 13. Representação esquemática do alinhamento realizado com o programa EST2GENOME para localização dos íntrons e éxons. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o primeiro grupo de ESTs de *Eucalyptus* ortólogos a esse gene (grupo *CesA7.2*). As setas indicam a localização e o sentido dos *primers* desenhados 44

Figura 14. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs denominado de EUC1, inferida pela utilização do programa EST2GENOME 44

Figura 15. Representação esquemática do alinhamento realizado com o programa EST2GENOME para localização dos íntrons e éxons. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o grupo de ESTs de *Eucalyptus* denominado de EUC1. As setas indicam a localização e o sentido dos *primers* desenhados 45

Figura 16. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs denominado de EUC2, inferida pela utilização do programa EST2GENOME 46

Figura 17. Representação esquemática do alinhamento realizado com o programa EST2GENOME para localização dos íntrons e éxons. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o grupo de ESTs de *Eucalyptus* denominado de EUC2. As setas indicam a localização e o sentido dos *primers* desenhados 48

Figura 18. Representação esquemática dos três *contigs* relativos ao gene denominado *CesA7.2*. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras coloridas representam os *contigs* de ESTs de *Eucalyptus* ortólogos a esse gene 49

Figura 19. Alinhamento e identificação de regiões intrônicas e exônicas para o gene *AtCesA8*, usando a seqüência genômica deste e um EST homólogo de *Eucalyptus*, através do programa EST2GENOME 50

Figura 20. Representação esquemática dos resultados obtidos pelo programa EST2GENOME para a localização das regiões correspondentes a éxons e íntrons no alinhamento dos ESTs ortólogos ao gene *AtCesA8*. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA4*, as barras azuis representam o *contig* de ESTs de *Eucalyptus* ortólogas a esse gene. As setas indicam a localização e o sentido dos *primers* desenhados 51

Figura 21. Confirmação da amplificação de fragmento do gene *CesA7.2* em superpools, utilizando os primers CesA72.3 e CesA72.4 (gel de agarose 1,5%) (1 – superpool 1; 17 – superpool 17; 21 – superpool 21 e + - controle positivo).....53

Figura 22. Confirmação da amplificação de fragmento do gene *CesA7.2* em indivíduos de diferentes espécies de *Eucalyptus* (1- *E. camaldulensis*; 2- *E. dunni*; 3- híbrido *E. urophylla* e *E. globulus*; 4- *E. grandis*), utilizando os primers CesA72.3 e CesA72.4 (gel de agarose 1,5%). 53

~~52~~

Figura 23. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC1.3 e EUC1.4 em clones da placa EUGR-BC-002-206 (gel de agarose 1,5%) Figura evidenciando posição relativa dos clones na placa.....55

Figura 24. Confirmação da amplificação de fragmentos pela utilização dos primers EUC1.3 e EUC1.4 em clones da placa 206 (gel de agarose 1,5%) (As bandas no gel se referem, da direita para a esquerda, aos clones da placa 206 1B, 1C, 1C, 2A, 2A, 2B e 3B, respectivamente).....56

Figura 25. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC2.3 e EUC2.4 nos *superpools* (gel de agarose 1,5%) (1 – *superpool* 1; 10 – *superpool* 10; 17 – *superpool* 17; 21 – *superpool* 21 e + - controle positivo). 57

Figura 26. Confirmação da amplificação de fragmentos pela utilização dos primers EUC2.3 e EUC2.4 em clones da placa 206 (gel de agarose 1,5%) Figura evidenciando a posição relativa dos clones na placa.....58

Figura 27. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC2.3 e EUC2.4 em indivíduos de diferentes espécies de *Eucalyptus* (1 e 2- *E. grandis;* 3- *E. urophylla*; 4- *E. dunni*; 5- *E. globulus*) (gel de agarose 1,5%) 59

Figura 28. Freqüência da ocorrência de clones positivos na biblioteca de BACs construída para *Eucalyptus grandis.*.....59

Figura 29. Fragmentos obtidos por amplificação de clones positivos, utilizando os *primers* CesA72.5 e CesA72.6. A penúltima canaleta (indicada pela seta) se refere ao controle positivo, seguida do *ladder 100pb* (Invitrogen Technologies).....60

Figura 30. Visualização dos produtos de PCR utilizando DNA dos clones selecionados e dois pares de *primers:* EUC2.3 e EUC2.4 (linha branca) e 72.3 e 72.4 (seta branca). As duas últimas canaletas são correspondentes aos marcadores de massa molecular (50ng e 100ng de DNA do fago λ, respectivamente).....61

Figura 31. *Contigs* obtidos pelo seqüênciamento parcial do gene *CesA7.2* de *Eucalyptus grandis.* Parte da seqüência depositada no *GenBank* como *CesA3* de *Eucalyptus grandis* é mostrada como referência.....62

Figura 32. Estrutura do gene AtCesA7. Os éxons são representados pelos retângulos e os íntron pelas linhas. Adaptado de Richmond (2000). 64

Figura 33. Alinhamento entre a seqüência *CesA* de *Eucalyptus grandis* e o contig formado pelo seqüênciamento de parte do gene *CesA7.2* de *Eucalyptus* realizado pelo programa Blast. 64

Figura 34. Porcentagem de *reads* detectados com elevado nível de similaridade ao gene *CesA7.2* em bibliotecas obtidas de xilema de diferentes espécies de *Eucalyptus* 66

Figura 35. Porcentagem de *reads* detectados com elevado nível de similaridade ao gene *CesA7.2* em bibliotecas de diferentes tecidos em diferentes espécies de *Eucalyptus:* XL (xilema); BC (DNA genômico); ML (folhas maduras); PU (folhas infectadas com *Puccina*); FX (biblioteca construída com floema de diferentes espécies) e SE (plântulas) 67

Figura 36. Alinhamento de seqüências genômicas de um fragmento do gene *CesA7.2* de diferentes espécies de *Eucalyptus* obtido pelo programa CLUSTALX. Os quadrados vermelhos indicam a localização dos SNPs 68

Figura 37. Alinhamento obtido pelo programa EST2GENOME evidenciando a localização do íntron. As setas indicam bases divergentes entre a seqüência do contig e a seqüência CesA3 de *Eucalyptus grandis* 68

Figura 38. Localização dos 6 SNPs encontrados pelo programa *Staden Package* na região flanqueada pelos *primers* CesA72.5 e CesA72.6 70

Figura 39. Localização dos 6 SNPs encontrados no fragmento seqüenciado pelos primers CesA72.5 e CesA72.6 71

Figura 40. Alinhamento obtido pelo programa *Staden Package* evidenciando a transversão que ocorre em apenas uma das espécies: *E. camaldulensis* (C1) 72

Figura 41. Eletroferogramas representativos da região abrangendo SNP na posição 105 72

Figura 42. Alinhamento evidenciando a presença de SNP na posição 423 do segundo éxon 73

Figura 44. Alinhamento evidenciando a ocorrência de prováveis SNPs na região intrônica 74

LISTA DE TABELAS

Tabela 01. Genes *CesA* de *Arabidopsis thaliana* utilizados como referência 19

Tabela 02. Genes *CesA* de outras espécies utilizados como referência..... 20

Tabela 03. Relação de ESTs de *Eucalyptus* identificados como relacionados a genes de celulose sintase de parede secundária..... 36

Tabela 04. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA4*..... 39

Tabela 05. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (*CesA7.1*) 42

Tabela 06. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (*CesA7.2*) 44

Tabela 07. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (fragmento EUC1)..... 46

Tabela 08. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (fragmento EUC2)..... 48

Tabela 09. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA8*..... 52

Tabela 10. Relação dos clones positivos para o fragmento do gene *CesA7.2*, identificados pela amplificação com os *primers* CesA72.3 e CesA72.4..... 54

Tabela 11. Relação dos clones positivos para o fragmento do gene *CesA7.2*, identificados pela amplificação com os *primers* EUC1.3 e EUC21.4 56

Tabela 12. Identificação dos clones positivos pela amplificação com os *primers* EUC2.3 e EUC2.4 58

Tabela 13. Distribuição de reads com elevado nível de similaridade aos *contigs* obtidos para o gene *CesA7.2* de *Eucalyptus grandis*, entre as bibliotacas do Projeto Genolyptus 66

Tabela 14. Caracterização do polimorfismo detectado em um fragmento do gene *CesA7.2* em diferentes espécies do gênero *Eucalyptus* 69

Tabela 15. Proporção de sítios polimórficos em regiões intrônicas e exônicas do gene *CesA7.2* em diferentes espécies de *Eucalyptus* 70

Tabela 16. caracterização dos SNPs detectados em relação ao tipo de substituição..... 71

LISTA DE SIGLAS

%	Porcentagem
ºC	Graus Celsius
ºS	Latitude sul
AtCesA	Gene celulose sintase de *Arabidopsis thaliana*
BAC	*Bacterial Artificial Chomosome*
BSA	Albumina de Soro Bovino
cDNA	DNA complementar
CesA	Gene celulose sintase
CESA	Proteína celulose sintase
CIA	Clorofórmio – álcool isoamílico
CTAB	*Cetyltrimethylammonium bromide*
DNA	Ácido desoxirribonucléico
dNTPs	Deoxirribonucleosídeo Trifosfato
EDTA	Ácido etileno diamono tetracético
EST	*Expressed sequence tag*
g	Gramas
H_2O	Água
Irx3	*Irregular xylem* (Mutação)
kb	Kilobase
KCl	Cloreto de potássio
kDa	Kilodaltons
KOAc	Acetato de potássio
M	Molar
m	Metros
mg	Miligramas
$MgCl_2$	Cloreto de magnésio
mL	Mililitros
mM	Milimolar
µg	Microgramas
µL	Microlitros
µm	Micrômetros
µM	Micromolar
NaCl	Cloreto de sódio
NaOH	Hidróxido de sódio
ng	Nanogramas
PCR	*Polimerase chain reaction*
pH	Potencial hidrogeniônico
pmol	Picomoles
RNAse	Ribonuclease
rpm	Rotações por minuto
SDS	Dodecil Sulfato de Sódio
SNP	*Single nucleotide polymorphism*
Taq	*Thermus aquaticus*
UDP	Uridina difosfato

ANOTAÇÃO E CARACTERIZAÇÃO PRELIMINAR DE GENES DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DE *Eucalyptus*

Autora: Marcela Mendes Salazar
Orientador: Prof. Dr. Alexandre Siqueira Guedes Coelho

Resumo

A celulose é o polímero mais abundante do planeta, sendo o principal constituinte da biomassa das plantas. Genes da família *CesA*, que codificam a enzima celulose sintase foram identificados em diversas espécies de plantas, especialmente em *Arabidopsis thaliana*, na qual três deles (*AtCesA4*, *AtCesA7* e AtCesA8) foram correlacionados com a produção de celulose na parede celular secundária. Espécies de *Eucalyptus* constituem um importante alvo de estudos de melhoramento genético, já que é o gênero florestal mais plantado no mundo, além de merecer destaque no mercado internacional de papel e celulose. A tecnologia de melhoramento molecular possibilita que a caracterização de genes de interesse possa ser aplicada a programas de melhoramento genético de espécies florestais visando o aumento da qualidade e produtividade dos seus produtos. Neste trabalho, 320 ESTs de *Eucalyptus*, divididos em quatro genes, responsáveis pela produção de parade celular secundária foram anotados, utilizando-se como referência os genes *AtCesA4*, *AtCesA7 e AtCesA8*. Para estes genes, *primers* foram desenhados a fim de que uma triagem em uma biblioteca de BACs de *Eucalyptus* fosse realizado. Uma maior ênfase foi dada ao gene ortólogo ao gene *AtCesA7* de *Arabidopsis thaliana* para o qual foi seqüenciado uma região de 1197 pares de base à patir da biblioteca de BACs e esta serviu de base para estudo do nível de expressão desse gene em diferentes espécies/tecidos, revelando que este é preferencialmente expresso em tecidos de xilema e fracamente expresso em folhas, o que corrobora com a atuação desses genes em paredes celulares secundárias. O nível de expressão desse gene foi maior em *E. urophylla* que em outras espécies estudadas. Seqüências de aproximadamente 500 pb também foram obtidas à partir de diferentes espécies de *Eucalyptus* e nestas, contendo um íntron flanqeado por dois

éxons, observou-se que a quantidade de SNPs identificados no íntron (4), como esperado, foi maior do que aquela encontrada nos éxons (1 em cada), embora o valor estimado para o índice de diversidade nucleotídica (π) relativo ao íntron (0,0824) tenha sido menor do que aquele estimado em um dos éxons (0,2029). Desse modo, espera-se que esse trabalho possa contribuir para uma melhor compreensão dos mecanismos envolvidos na biossíntese e regulação da via de celulose em espécies de *Eucalyptus*, bem como subsidiar estudos de mapeamento genético e análise de desequilíbrio de ligação nesse gênero.

ANNOTATION AND PRELIMINARY CHARACTERIZATION OF CELLULOSE SINTHASE GENES IN DIFFERENT *Eucalyptus* SPECIES

Author: Marcela Mendes Salazar
Adviser: Prof. Dr. Alexandre Siqueira Guedes Coelho

Abstract

Cellulose is the world's most abundant polymer, being the main constituent of plant biomass. Genes from *CesA* family, that encodes the cellulose synthase enzyme, was identified in several plant species, especially in *Arabidopsis thaliana*, in which three of them (*AtCesA4, AtCesA7* e *AtCesA8*) were associated with secondary cell wall cellulose production. *Eucalyptus* species represent an important target of genetic improvement studies, since it is the most planted forest genus in the world, beyond deserving a special place in the international market of cellulose and paper. The molecular breeding technology enables that gene characterization can be applied to forest species genetic improvement programs with the aim to improve the quality and productivity of its products. In this work, 320 *Eucalyptus* ESTs, separated in four genes related with cellulose production in secondary cell wall, using *AtCesA4, AtCesA7* e *AtCesA8* genes as reference. For these genes, primers were designed in order to screen an *Eucalyptus* BACs library. An emphasis was given to genes that are orthologs to *AtCesA7* from *Arabidopsis thaliana* for witch it was sequenced a 1197 base pairs region from the BACs library and this served as a support to expression level studies of this genes in different species/tissues, showing that this is preferentially expressed in xylem and weakly expressed in leaves. The expression level of this gene was higher in *E. urophylla* than in other species studied. Sequences from approximately 500bp was obtained from different *Eucalyptus* species and in these, with one intron between two exons, the amount of SNPs in the intron (4), as waited, was higher than that found in exons (1 in each), although the intron nucleotide diversity index (π) (0,0824) were less than in the exon (0,2029). In this manner, one expects that this work can contribute for one better understanding of the mechanisms involved in biosynthesis and

regulation of the cellulose pathway in *Eucalyptus* species, as well as subsidize genetic mapping studies and linkage disequilibrium analysis for this genus.

1 INTRODUÇÃO

A celulose é o polímero mais abundante do planeta, sendo o principal constituinte da parede celular da maioria das células vegetais maduras (Dhugga, 2001). Apesar da sua grande importância, as informações sobre os mecanismos moleculares envolvidos na sua biossíntese, bem como sobre a regulação deste processo, são recentes e ainda elementares. Genes da família *CesA* (celulose sintase) foram identificados recentemente em diversas espécies de plantas (*Arabidopsis thaliana* e *Populus tremuloides*,) (Samuga & Joshi, 2002; Taylor *et al*., 1999). A função destes genes foi atribuída à codificação da subunidade catalítica do complexo celulose sintase, um complexo enzimático localizado na membrana plasmática e responsável pela síntese de celulose (Doblin *et al.*, 2002). Em *A. thaliana*, pelo menos dez isoformas desse gene foram identificadas, sendo que três delas: *AtCesA4*, *AtCesA7* e *AtCesA8* são responsáveis pela síntese de celulose na parede celular secundária (Joshi *et al.*, 2004). Uma mutação no gene *AtCesA7* (*irx3*) foi correlacionada com a diminuição da quantidade total de celulose produzida pela planta (Taylor *et al*., 1999).

A utilização de espécies do gênero *Eucalyptus* como matéria-prima para fabricação de celulose fez do *Eucalyptus*, em meados de 1950, o gênero florestal mais plantado no mundo. Dessa maneira, países produtores de celulose de *Eucalyptus*, como o Brasil, assumiram um papel de destaque no mercado internacional de celulose (Bracelpa, 2005; Lima, 1996).

O gênero *Eucalyptus* é constituído por mais de 670 espécies, além de numerosas subespécies, variedades e híbridos. O centro de origem do gênero é o continente australiano. As espécies já identificadas exibem grande variação estrutural. O gênero inclui espécies com hábitos desde arbustivo até arbóreo, com indivíduos que atingem até 75 metros de altura, como é o caso de indivíduos da espécie *Eucalyptus grandis* (Euclid, 2006; Paula, 2005; IPEF - chave de identificação, 2006). A introdução do gênero no Brasil se deu por volta de 1868, inicialmente para fins ornamentais, tendo a sua utilização industrial sido inicializada somente por volta de 1910, em decorrência da implantação da Companhia Pau-

lista de Estradas de Ferro que objetivava a produção de lenha para locomotivas (Lima, 1996).

O *Eucalyptus*, hoje, é plantado em diversos países, sendo considerado no Brasil a principal cultura florestal. Devido à grande adaptação de diversas espécies a diferentes países, muitas delas são utilizadas industrialmente como: *Eucalyptus grandis*, *E. saligna*, *E. urophylla*, *E. globulus* e híbridos de *E. grandis* e *E. urophylla* (Bertola, 2005). No Brasil, o cultivo de *Eucalyptus* merece um destaque principalmente no que se refere à sua utilização em programas de reflorestamento e produção de papel e celulose. Em 2002, o *Eucalyptus* representava 69% da área reflorestada do país. O setor de produção de papel e celulose do país situava-se, neste ano, entre os três maiores fornecedores do mercado mundial (Gonzáles *et al.*, 2002). Já no ano de 2005 a área ocupada por espécies de *Eucalyptus* atingiu um total de 75% da área total destinada à plantação de árvores para indústrias de papel e celulose. Os outros 24% eram ocupados por *Pinus* e 1% por outras espécies. Estima-se que cerca de 2,6 milhões de hectares de florestas nativas foram poupados da destruição. Nesse mesmo ano, o setor de papel e celulose do país gerou um saldo comercial de 2,5 bilhões de dólares e o Brasil alcançou o lugar de maior produtor mundial de celulose de fibra curta. Os investimentos realizados no setor nos últimos anos têm possibilitado às empresas atingir padrões internacionais no que se refere, tanto à qualidade dos produtos, quanto à produtividade e proteção ao ambiente (Bracelpa, 2005b).

É importante considerar, no entanto, que a manutenção da competitividade dos produtos de *Eucalyptus* no mercado internacional exige, principalmente, que investimentos em pesquisa sejam canalizados para a solução de problemas existentes nas diferentes etapas do processo produtivo. Investimentos neste sentido vêm sendo realizados nos últimos anos pelo setor florestal, no intuito de se conseguir um aumento nos ganhos de produtividade de celulose, através das práticas de melhoramento genético. Neste contexto, destaca-se também o Projeto Genolyptus – Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*, fruto do esforço conjunto de diversas instituições públicas de pesquisa (Embrapa e universidades) e diversas empresas da área florestal, visando a identificação e caracterização de genes relacionados à melhoria da qualidade e produtividade dos derivados da madeira de *Eucalyptus* (Gonzáles *et al.*, 2002; Revista Madeira, 2003, PROJETO GENOLYPTUS – Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*).

A crescente disponibilidade das modernas técnicas de análise molecular têm permitido o desenvolvimento de diversas tecnologias voltadas para a aplicação da análise

genômica a programas de melhoramento genético, caracterizando uma abordagem conhecida como melhoramento molecular (*molecular breeding*). Neste novo cenário, a caracterização de genes de interesse, de suas regiões intrônicas e exônicas, dos seus possíveis sítios de regulação, além da detecção de polimorfismos e da avaliação da magnitude da diversidade nucleotídica presente em diferentes populações e espécies, são estudos primários de suma importância. Espera-se que os resultados destes estudos possam contribuir efetivamente para a melhor compreensão dos mecanismos envolvidos na biossíntese de produtos de interesse, bem como dos aspectos relacionados à regulação destes processos. Tal entendimento deverá possibilitar, em última instância, o desenvolvimento e a aplicação de estratégias de melhoramento genético molecular visando o aumento da qualidade e produtividade dos produtos de origem florestal.

Este trabalho foi desenvolvido com o objetivo de se realizar uma caracterização dos genes de celulose sintase em espécies do gênero *Eucalyptus*, com ênfase na análise do gene homólogo ao gene *CesA7* de *Arabidopsis thaliana*. No intuito de subsidiar futuras ações de mapeamento genético e de análise de desequilíbrio de ligação, buscou-se também realizar a detecção de sítios polimórficos (SNPs) e a avaliação do nível de diversidade molecular presente neste gene, entre diferentes espécies do gênero *Eucalyptus*.

2 OBJETIVOS

i. Identificação e classificação preliminar dos ESTs de genes de celulose sintase obtidos no âmbito do Projeto Genolyptus, utilizando como referência a classificação dos genes de celulose sintase descritos para *Arabidopsis thaliana;*

ii. Análise preliminar das seqüências consenso obtidas para os diferentes genes de celulose sintase em *Eucalyptus* e desenho de *primers* específicos para cada gene;

iii. Identificação de clones da biblioteca BACs, construída no âmbito do Projeto Genolyptus, contendo genes de celulose sintase;

iv. Caracterização de um gene de celulose sintase de *Eucalyptus*, homólogo ao gene *CesA7* (celulose sintase de parede secundária) de *Arabidopsis thaliana*, em termos de seqüência, presença de sítios polimórficos, magnitude do polimorfismo inter-específico e nível de expressão em diferentes espécies/tecidos.

3 REVISÃO DE LITERATURA

3.1 O GÊNERO *Eucalyptus*

O *Eucalyptus* é uma angiosperma pertencente à família *Myrtaceae*. O centro de origem do gênero é a Oceania, onde ocorre em cerca de 90% de toda a área da Austrália. Mais de 670 espécies, numerosas subespécies, variedades botânicas e híbridos já foram identificados, sendo que duas espécies (*Eucalyptus urophylla* e *E. deglupta*) possuem ocorrência natural fora da Oceania (Bertola, 2005; Dickinson, 1969).

Dentre as muitas espécies já identificadas, existe uma grande variação no que diz respeito à biologia e morfologia das árvores. Quanto à forma de crescimento, estas podem possuir troncos bifurcados ou troncos únicos. As primeiras alcançando até 10 metros de altura e as com troncos únicos podendo atingir até os 75 metros de altura, como ocorre por exemplo em indivíduos de *Eucalyptus grandis* (Euclid, 2006; IPEF - chave de identificação, 2006).

O córtex apresenta uma característica marcante nas espécies de *Eucalyptus*. Como há um contínuo crescimento interno deste nas árvores, a camada mais externa morre a cada ano. Em algumas espécies, essa casca se desprende completamente, expondo assim uma nova camada de córtex vivo. Tais espécies são designadas como possuidoras de casca lisa. Em contraste, existem várias espécies com tronco áspero. Neste, o córtex, morto, seca e não há o desprendimento do mesmo (Euclid, 2006).

A maioria das folhas adultas de *Eucalyptus* possui uma forma lanceolada ou falciforme, possuindo cores semelhantes em ambos os lados, sendo geralmente de coloração verde forte. A presença de glândulas de óleo nas folhas é outra característica notável dentro das espécies de *Eucalyptus*. As flores de todas as espécies são hermafroditas, agrupadas em corimbos ou umbelas, sendo a maioria de cor branca (Euclid, 2006).

Em relação à biologia reprodutiva, as espécies de *Eucalyptus* utilizam a alogamia como principal sistema reprodutivo (Pryor, 1976, citado por Paula, 2005), porém, até aproximadamente 30% de autogamia pode ocorrer, condição típica de espécies que apre-

sentam um sistema misto de reprodução, com níveis intermediários de endogamia. A elevada taxa de alogamia se deve à existência de protandria, ou seja, o amadurecimento e liberação dos grão de pólen ocorrem quando os estigmas ainda se encontram imaturos e não receptivos. A taxa de 30% de autopolinização ocorre devido ao fato de uma mesma planta apresentar flores com diferentes estágios de maturação (Eldridge, 1978 citado por Paula, 2005).

Acredita-se que o gênero *Eucalyptus* tenha sido introduzido na Europa por volta de 1774, tendo sido descrito pela primeira vez pelo botânico francês L'Héritier de Brutelle, e só posteriormente alcançado outros países. Na América do Sul, foi introduzido inicialmente no Chile em 1823. No Brasil, acredita-se que tenha sido introduzido em 1868, embora a determinação desta data ainda seja discutida, tendo sido utilizado como uma planta decorativa por um longo período. A expansão da cultura de *Eucalyptus* para fins industriais só teve início graças à Companhia Paulista de Estradas de Ferro, por volta de 1910, visando inicialmente a produção de lenha para locomotivas e batentes (Bertola, 2005; Lima, 1996). Sua utilização como matéria-prima para produção de celulose teve início somente nos anos 1950. O *Eucalyptus* atingiu uma posição de destaque no mercado brasileiro em 1980, quando o país atingiu a primeira colocação como produtor e exportador de celulose branqueada (Lima, 1996).

O *Eucalyptus*, hoje, é cultivado em diversos países de clima temperado e tropical, ocupando regiões de grande diversidade ambiental. O *Eucalyptus* pode ser encontrado em habitats variados, tanto em termos de precipitação, quanto de temperaturas, ocorrendo ainda em regiões tropicais equatoriais até regiões semi-áridas, e em elevações que vão desde o nível do mar até mais ou menos 4.000 metros de altitude. Algumas espécies (*E. robusta* e *E. camaldulensis*) são capazes de crescer também em regiões alagadas. Tal versatilidade do gênero possibilitou a expansão de sua produção em diferentes regiões do território brasileiro, bem como em todo o mundo (Lima, 1996).

O cultivo do *Eucalyptus* se tornou um grande aliado no que diz respeito à conservação da natureza, uma vez que, devido a características como fácil adaptação às diferentes condições edafo-climáticas e rápido crescimento, tornou-se uma alternativa interessante para programas de reflorestamento e contra as devastações das florestas naturais (Bracelpa, 2005a). Assim, as plantações florestais (formações de florestas a partir de regeneração artificial) são uma alternativa para o suprimento do consumo de madeira para fins industriais e para geração de energia de um modo mais racional quando comparadas às florestas naturais. As últimas já estão escassas e ainda ocupam áreas onde sua presença é

muito importante para fins de conservação ambiental (Lima, 1996). Devido a essas vantagens, diversas espécies do gênero foram eleitas como padrão para a produção de matéria-prima florestal para diferentes usos, tornando o *Eucalyptus* o gênero florestal mais plantado no mundo.

A matéria-prima obtida através das árvores de *Eucalyptus* é utilizada em diferentes setores industriais. A fibra da madeira é utilizada para produção de celulose. A madeira é utilizada para produção de móveis e na construção civil. O óleo de *Eucalyptus* é utilizado pela indústria alimentícia, farmacêutica, além de ser usado para fabricação de produtos de limpeza e cosméticos. O pólen de suas flores é ainda utilizado na produção de mel (Bertola, 2005).

Dentre as espécies de *Eucalyptus* mais plantadas no Brasil e no mundo, algumas merecem destaque: *Eucalyptus grandis*, *E. saligna*, *E. urophylla*, *E. globulus* e híbridos de *E. grandis* e *E. urophylla* (Bertola, 2005). Em 2002, o *Eucalyptus* representava 69% da área reflorestada do país, ou seja, cerca de 1,5 milhões de hectares. O setor de produção de papel e celulose situava-se entre os três maiores fornecedores do mercado (Gonzáles *et al.*, 2002).

Em 2005 a área destinada à plantação de árvores para as indústrias de papel e celulose atingiu 1,7 milhões de hectares, sendo 75% ocupados por árvores de *Eucalyptus*, 24% por *Pinus* e outras espécies ocupando apenas 1%. Estima-se que cerca de 2,6 milhões de hectares de florestas nativas foram poupados da destruição. Além disso, o setor de papel e celulose brasileiro gerou um saldo comercial para o país de 2,5 bilhões de dólares. Desse modo, o Brasil conseguiu alcançar o lugar de maior produtor mundial de celulose de fibra curta em 2005, chegando a produzir 6 milhões de toneladas por ano. Os investimentos realizados no setor nos últimos anos possibilitaram às empresas atingir padrões internacionais no que se refere tanto à qualidade dos produtos, quanto em produtividade e proteção ao ambiente (Bracelpa, 2005b).

3.1.1 Caracterização de algumas espécies do gênero *Eucalyptus*

a) *Eucalyptus grandis* Hill. ex Maiden.

As árvores desta espécie são muito altas (40 a 45m), podendo eventualmente atingir 75m de altura. Possui tronco grosso (geralmente 2m de diâmetro), com casca de textura lisa (IPEF - chave de identificação, 2006).

Ocorre naturalmente na Austrália, ao norte do Estado de New South Wales, ao sul de Queensland (próximo à região costeira e na parte central), e ao norte de Queensland em área de altitude (300 a 900m). A madeira de *E. grandis* é leve e fácil de ser trabalhada. É utilizada intensivamente, na Austrália e na República Sul Africana, como madeira de construção. Normalmente, a madeira oriunda de árvores com rápido crescimento apresenta problemas de empenamento, contrações e rachaduras. Plantações convenientemente manejadas, no entanto, podem produzir madeira excelente para serraria e laminação. Essa espécie é a principal fonte de matéria prima para celulose e papel do Estado de São Paulo (Ferreira, 1979) e é provavelmente a espécie mais plantada para a produção de madeira para fins industriais no mundo, sendo a maior área de plantio concentrada no Brasil e na África do Sul (FAO, 2006).

b) *Eucalyptus urophylla* S. T. Blake

Espécie pertencente ao subgênero *Symphyomyrthus* (Pryor & Johson, 1971, citado por Laerte, 2002), sua área de ocorrência natural situa-se em Timor e outras ilhas a leste do arquipélago indonesiano, entre as latitudes de 8 a 10°S e altitudes de 400 a 3.000m. As árvores de *E. urophylla* apresentam um porte elevado (de 30 a 60m de altura), forte dominância apical, possui casca rugosa, podendo ser lisa em função da altitude e de cor marrom a quase preta. As folhas das plantas adultas são alongadas, estreitas e pecioladas. Os frutos podem estar em duas formas distintas: campanulada ou hemisférica, em altitudes superiores a 1000m, e cônico em altitudes inferiores a 1000m (Martin & Cossalter, 1976, citados por Pigato & Lopes, 2001).

Na área de ocorrência natural da espécie a madeira é utilizada para construções e estruturas que demandem alta resistência. Entre as espécies de *Eucalyptus* plantadas no Brasil, *E. urophylla* é a mais plantada e a que apresenta maior potencial de crescimento em relação à resistência do fungo causador do cancro do *Eucalyptus* (*Cryphonectria cubensis*). É utilizada para fabricação de celulose e papel, serraria, produção de carvão e outros (Scanavaca Junior, 2002; FAO, 2006).

c) *Eucalyptus globulus* Labill.

É uma árvore de médio a grande porte, 15 a 25m, podendo alcançar 70m de altura (Santos, 2005). É nativa da Tasmânia e sudeste da Austrália. Foi a primeira espécie a se tornar mais largamente conhecida fora da Austrália, sendo sua madeira usada para fabricação de papel, lenha e construção civil. No sul da Austrália esta espécie é preferida devido à sua alta taxa de crescimento, pequeno período de rotação e qualidades ótimas da madeira (FAO, 2006).

3.2 A PAREDE CELULAR DAS CÉLULAS VEGETAIS

3.2.1 Estrutura da parede celular

Uma das maiores distinções entre células animais e células vegetais é a parede celular. Esta estrutura também é sintetizada por alguns protistas, algas, bactérias (*Acetobacter*, *Agrobacterium* e *Rhizobium*) e está presente na maioria das células de fungos. A parede celular fornece, principalmente, sustentação para estes organismos (Salisbury & Ross, 1991; Richmond, 2000). Nas plantas, este órgão desempenha importantes funções. A parede celular controla o volume da célula, pois impede que a membrana celular se rompa quando ocorre intensa entrada de água na célula. Possui enzimas que participam da absorção e transporte de nutrientes e secreção de resíduos, atua na defesa do organismo contra patógenos, e certos hormônios presentes na parede também ajudam a regular o desenvolvimento e crescimento da planta (Raven *et al.*, 2001; Taiz & Zeiger, 1998).

A parede celular das plantas (Figura 1) é composta de três camadas principais: parede primária, parede secundária e lamela média. Esta composição varia de acordo com o tipo celular e a idade da planta, principalmente no que se refere à presença ou ausência da parede celular secundária. Células em crescimento, de parênquima, células fotossintetizantes presentes nas folhas e alguns outros tipos celulares, possuem somente uma parede celular primária. Esta se caracteriza por ser fina (em torno de 1 a 3 μm), ser constituída de 9 a 25% de celulose, possuindo também hemicelulose e algumas proteínas (Raven *et al.*, 2001; Salisbury & Ross, 1991). A água também está presente em abundância na parede primária e é indispensável para a manutenção das propriedades físicas da parede, como por exemplo, a rigidez. As microfibrilas de celulose estão dispersas em uma matriz amorfa cuja constituição é predominantemente polissacarídica (hemicelulose e pectina), possuindo também algumas proteínas estruturais (Taiz & Zeiger, 1998).

Já na maioria de células maduras, especialmente naquelas presentes em tecidos de xilema, existe uma parede celular secundária que fornece o suporte para a planta. Esse suporte mecânico se dá, principalmente, devido à presença de fibras de constituição celulósica compactadas e lignina (Taiz & Zeiger, 1998). As diferenças nas propriedades físicas de paredes primárias e secundárias se devem, em grande parte, à quantidade de cadeias de celulose que estão presentes em cada microfibrila. Em paredes secundárias cada microfibrila possui um número muito maior de cadeias de celulose quando comparadas com o número destas presente nas microfibrilas constituintes da parede celular primária (Taylor *et al.*, 1999). A parede celular de células presentes na madeira é composta por diversas camadas de parede celular secundária consistindo de uma rede de celulose e hemicelulose saturada com lignina (Andersson-Gunneras *et al.*, 2006).

A lamela média é uma camada fina, constituída predominantemente de pectinas (polissacarídeo formado por moléculas de ácido galacturônico). Está situada na região de contato entre as paredes primárias de duas células adjacentes e está repleta de plasmodesmas, os quais são responsáveis pela comunicação e transporte de nutrientes célula-a-célula (Taiz & Zeiger, 1998). A natureza das proteínas estruturais, que constituem apenas uma pequena porção da parede celular em crescimento, e suas funções precisas, ainda permanecem indeterminadas (Dhugga, 2001).

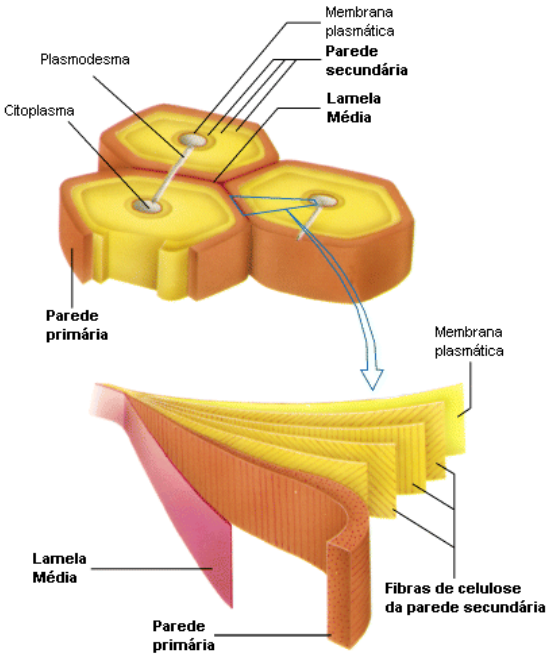


Figura 01. Estrutura da parede celular de plantas, evidenciando as diversas camadas que formam a parede celular secundária. Adaptado de Lodish *et al.* (2000).

3.2.2 Estrutura e Biossíntese de Celulose

O principal constituinte da parede celular é a celulose, que consiste em um polissacarídeo hidrofóbico formado pela polimerização de resíduos de glucose, a partir de um substrato de UDP-glucose, formando um homopolímero β-1,4 glucano (Burton *et al.*, 2005). Nas paredes celulares primárias das plantas, a celulose contribui com cerca de 20 a 30% da massa seca (Mellerowicz, 2001, citado por Samuga & Joshi, 2004). As microfibrilas de celulose das paredes primárias controlam o tamanho e a forma das células vegetais. Quando comparadas às paredes celulares primárias, paredes celulares secundárias contêm celulose com alto grau de polimerização fornecendo força mecânica e rigidez para toda a planta (Samuga & Joshi, 2004).

O processo de biossíntese de celulose é um dos mais importantes da natureza, já que esse polissacarídeo é o principal constituinte das paredes da maioria das células maduras (Dhugga, 2001). Porém, em contraste com a grande abundância e importância da celulose, pouco se sabe sobre a regulação dos mecanismos moleculares envolvidos na sua biossíntese em plantas. Sabe-se, entretanto, que a biossíntese desse polímero é um processo complexo e que exige altos níveis de organização. A cadeia formada pelos resíduos de glucose é linear, pois durante a ligação um dos resíduos sofre uma rotação de 180º em relação ao resíduo anterior. Esta forma linear que a celulose assume permite que haja uma grande interação entre cadeias formando uma estrutura rígida. Pontes de hidrogênio se formam dentro das cadeias e entre elas, de modo que uma estrutura cristalina pode ser observada (Delmer & Amor, 1995). Toda essa interação entre as cadeias garante a essa estrutura propriedades notáveis, como: alta resistência à tensão, insolubilidade e a capacidade de ser relativamente imune à degradação enzimática e química (Taiz & Zeiger, 1998).

A estrutura que se forma pela interação íntima entre cadeias de celulose é chamada de microfibrila de celulose. Esta estrutura pode conter de 36 a 200 cadeias, uma variação que ocorre entre espécies. O comprimento das microfibrilas também é bastante variável. A organização das cadeias de celulose também se dá em níveis mais altos. Um conjunto de microfibrilas é chamado de macrofibrilas. A reunião deste último forma um feixe de celulose (Figura 02) (Delmer & Amor, 1995; Taiz & Zeiger, 1998). As microfibrilas de celulose são sintetizadas por um grande complexo protéico (>500 kDa) localizado na membrana celular. Esse complexo é composto por várias subunidades de uma enzima de-

nominada celulose sintase, formando o complexo celulose sintase, que catalisa a transferência de glucose para a cadeia glucano crescente, utilizando como substrato uridina-difosfo-glucose (UDP-glucose) (Doblin *et al.*, 2002).

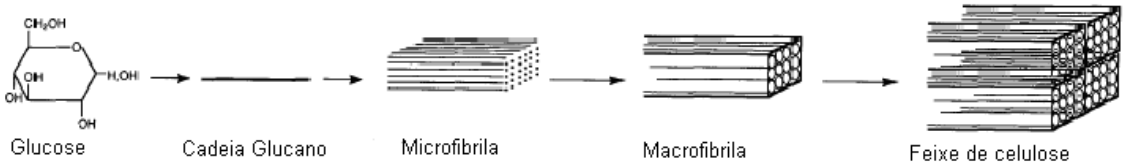


Figura 02. Estrutura da interação das cadeias de celulose. Adaptado de Delmer & Amor (2000).

3.2.3 Genes *CesA*

Os genes *CesA* são responsáveis pela codificação das subunidades catalíticas do complexo celulose sintase em diversos organismos (Doblin *et al.*, 2002). Nas últimas décadas, diversas técnicas bioquímicas foram utilizadas na tentativa de se identificar as "polissacarídeos sintases" de plantas. Tal procedimento, neste caso, não obteve o mesmo sucesso obtido em bactérias. Nesses microrganismos, o gene da celulose sintase (*CesA*) foi isolado através da purificação enzimática e seqüenciamento de peptídeos. Foi somente com a modernização de técnicas de análise genômica que estes genes foram identificados em plantas. Em 1996, clones de cDNA de genes responsáveis pela produção de celulose em fibras de algodão em desenvolvimento foram isolados e observou-se que as proteínas preditas a partir do cDNA continham regiões de aminoácidos que eram altamente similares às regiões da proteína CESA bacteriana (Dhugga, 2001). A partir de então, muitos genes com seqüências similares a *CesA* de algodão foram isolados de outras espécies de plantas. Em *Arabidopsis thaliana*, devido à disponibilidade de um genoma totalmente seqüenciado, foi possível identificar 10 isoformas pertencentes a uma grande família destes genes. Estas isoformas foram nomeadas de *AtCesA1* a *AtCesA10* (Samuga & Joshi, 2002). A análise com mutantes foi realizada na tentativa de se identificar a função destes genes. O sucesso em tal identificação foi conseguido para a maioria destes genes e revelou que três genes de *A. thaliana* (*AtCesA1*, *AtCesA3* e *AtCesA6*) parecem estar envolvidos na formação da parede celular primária, enquanto outros três genes (*AtCesA4*, *AtCesA7* e *AtCesA8*) estão relacionados ao desenvolvimento da parede celular secundária (Joshi *et al.*, 2004). Devido à

grande importância econômica da celulose de árvores para indústrias e produtos florestais (indústria de papel e celulose) muitos trabalhos estão sendo realizados no intuito de se identificar os genes de celulose sintase em diversas espécies florestais como *Populus tremuloides*, *Pinus pinaster*, entre outras (*PtrCesA1* e *PtrCesA2*, por exemplo) (Samuga & Joshi, 2002; 2004).

O complexo celulose sintase (Figura 3) é assim chamado por ser formado por diversas subunidades de uma enzima. É um complexo integral de membrana (contendo regiões hidrofóbicas e hidrofílicas), sendo a região catalítica voltada para o citoplasma, e está estruturado sob a forma de uma roseta contendo seis lobos distintos, cada lobo contendo 6 subunidades catalíticas da enzima, ou seja, esta estrutura contém, provavelmente, 36 proteínas CESA (Doblin *et al.*, 2002, Gardiner *et al.*, 2003; Brown Jr. & Saxena, 2000).

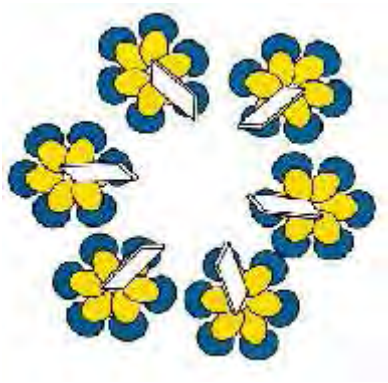


Figura 3. Estrutura em roseta do complexo celulose sintase. Os retângulos brancos representam cadeias de celulose recém-formadas. Adaptado de Brown Jr. & Saxena (2000).

A subunidade catalítica se liga ao substrato UDP-glucose e os adiciona à cadeia de celulose nascente que é liberada no espaço extracitoplasmático. Assim, cada proteína CESA produz uma cadeia de celulose que se une à cadeia adjacente na roseta, formando uma microfibrila (Gardiner *et al.*, 2003; Brown Jr. & Saxena, 2000).

No mínimo três genes *CesA* são requeridos para a síntese de celulose nas paredes secundárias de tecidos de xilema em desenvolvimento (*CesA4*, *CesA7* e *CesA8*) (Delmer & Amor, 1995; Gardiner *et al.*, 2003). A função desses genes foi desvendada através de seus mutantes. Mutações nos genes *AtCesA4*, *AtCesA7* e *AtCesA8* (descritas como *irx5*, *irx3* e *irx1*, respectivamente) causam diminuição na quantidade total de celulose nas paredes celulares secundárias. Essa mutação *irx* (*irregular xylem* - xilema irregular) é caracte-

rizada por um colapso em vasos de xilema que ocorre pela falta de resistência à pressão exercida durante o transporte de água. Isso ocorre porque a formação de parede celular é anormal e o caule em formação é muito fraco (Taylor *et al.*, 1999). Esses três genes são responsáveis pela codificação de subunidades catalíticas diferentes do complexo celulose sintase, mas todos são essenciais para a formação da celulose na parede celular secundária de uma mesma célula. Todas as subunidades são igualmente importantes durante o processo de montagem do complexo celulose sintase (Taylor *et al.*, 2003). Apesar de todos esses genes serem expressos em uma mesma célula, uma mutação em apenas um deles é suficiente para reduzir consideravelmente a quantidade de celulose depositada nas paredes celulares secundárias. Esse fato indica que a mutação de um desses genes não é suprimida por outros genes (Zhong *et al.*, 2003).

Os genes *CesA4*, *CesA7* e *CesA8* são expressos em grande quantidade no tronco, e fracamente expressos em outros tecidos corroborando com a informação genética de que esses genes codificam componentes do complexo celulose sintase que interferem na síntese da parede celular secundária (Hamann *et al.*, 2004). Taylor *et al.* (1999) constataram que o gene *CesA7* não é expresso em tecidos foliares em *Arabidopsis thaliana*. Trabalhos recentes demonstraram que genes de celulose sintase são fracamente expressos nas folhas e além disso o gene *CesA7* é consistentemente expresso em níveis mais altos que o gene *CesA8*, porém eles exibem o mesmo perfil de expressão sugerindo que esses genes são co-regulados (Brown *et al.*, 2005, Suzuki *et al.*, 2006).

3.2.4 Genes *AtCesA7*

O gene *AtCesA7* de *Arabidopsis thaliana* está relacionado com a formação de celulose na parede celular secundária. A caracterização do mutante *irx3* (*irregular xylem 3*), carregando uma mutação sem sentido neste gene, demonstrou que a ausência do produto funcional deste gene foi correlacionada com uma diminuição significativa, de aproximadamente 70%, na quantidade total de celulose produzida pela planta e por isso a flexibilidade do caule das plantas se torna maior, de maneira que as torna incapazes de progredir com o crescimento perpendicular ao solo (Taylor *et al.*, 1999). A análise da composição celular do mutante *irx3* revela que existe uma diferença significativa em relação ao tipo selvagem, decorrente da diminuição da quantidade de celulose no caule. O mutante contém aproximadamente um terço da quantidade de celulose (Taylor *et al.*, 2004). Além disso, o

tipo selvagem é caracterizado por elementos de xilema abertos e com forma arredondada, enquanto que o mutante possui vasos colapsados, geralmente oclusos e de forma irregular, como mostrado na Figura 4 (Brown *et al*., 2005).

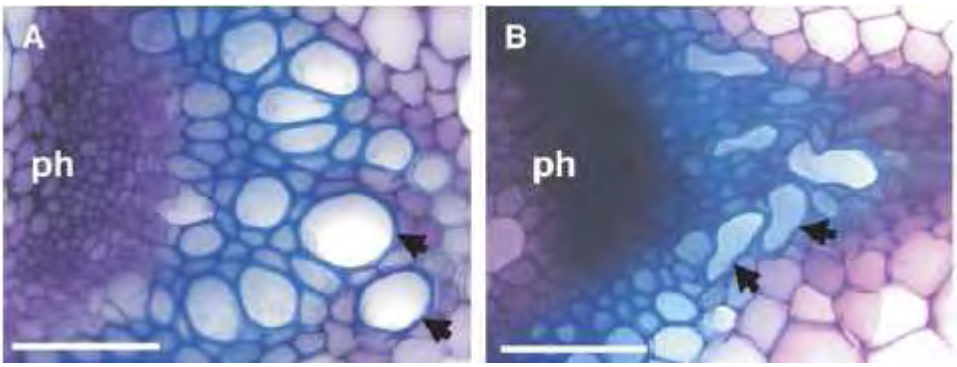


Figura 4. Secção tranversal do sistema vascular de *A. thaliana* evidenciando elementos do xilema no tipo selvagem (A) e no mutante *irx3* (B) (ph – floema). Adaptada de Brown *et al*. (2005).

O gene *AtCesA7* mostrou uma identidade de 85% com o gene *PtrCesA2*, identificado em *Populus tremuloides* (Samuga & Joshi, 2002) e de 88% com os genes *PtCesA7* e *PtCesA17* de *Populus trichocarpa* (Suzuki *et al*., 2006).

3.2.5 Polimorfismos de um único nucleotídeo (SNPs)

Determinadas regiões do DNA apresentam polimorfismos em um único nucleotídeo de uma seqüência. Esse polimorfismo é definido como um SNP (polimorfismo em nucleotídeos únicos). O termo SNP (*Single Nucleotide Polimorfism*) se refere, mais restritamente, à alternância que ocorre entre os quatro nucleotídeos em um dado sítio de uma seqüência, em uma dada população. Em outras palavras, SNPs são diferenças nas seqüências de diferentes alelos de um dado lóco que ocorrem em um único par de bases. Nessa definição não estão incluídas as inserções e as deleções. Considera-se geralmente, como necessário, que o alelo mais raro desse lóco polimórfico possua uma freqüência alélica superior ou igual a 1% (Guimarães & Costa, 2002; Rafalski, 2002). Definições mais abrangentes de SNPs incluem polimorfismos decorrentes de inserções e/ou deleções em seqüências genômicas e variantes com freqüência alélica menor que 1% (Brookes, 1999).

Na verdade, os SNPs podem ser polimorfismos bi, tri ou tetra-alélicos, sendo os bi-alélicos os mais comuns (Brookes, 1999). Podem ocorrer tanto dentro de éxons, quanto em íntrons. O polimorfismo de nucleotídeos únicos que ocorrem dentro de regiões exônicas pode causar modificações na seqüência de aminoácidos que podem ser do tipo conservativa, quando o aminoácido substituído possui as mesmas características físico-químicas do aminoácido original, ou pode resultar em uma substituição não conservativa, quando os aminoácidos envolvidos no polimorfismo possuem características físico-químicas diferentes. A substituição dos nucleotídeos pode ainda ocorrer de maneira a gerar o mesmo aminoácido, já que o código genético é degenerado, sendo estas substituições chamadas de silenciosas (Guimarães & Costa, 2002; Lehninger *et al.*, 1992).

Os dados gerados a partir de seqüenciamento, principalmente a identificação de SNPs, podem ser explorados no campo da genética de populações. Os SNPs refletem uma mutação que ocorreu no passado e os indivíduos que apresentam essa variação estão marcados com essa herança. Através do perfil exibido pela variação dos SNPs, podemos inferir os genes ancestrais desses alelos. Muitos SNPs estão diretamente relacionados a doenças humanas ou a genes de interesse econômico. Mas em muitos casos, em que a associação não é direta, eles funcionam como marcadores moleculares para localizar esses genes através de estudos de herança (Stoneking, 2001; Kwok & Gu, 1999). Neste contexto, a diversidade nucleotídica é um forte indicador biológico da história evolutiva que afeta os indivíduos (Chakravarti, 2001).

O alto nível de polimorfismo dentro das muitas espécies de plantas facilita a identificação dos SNPs, sendo o seqüenciamento de produtos de PCR de um certo fragmento de DNA o meio mais direto de se identificar estes polimorfismos. O seqüenciamento realizado em regiões intrônicas aumenta em aproximadamente três vezes a freqüência de polimorfismo encontrada dentro de uma seqüência de DNA. Os SNPs são o tipo de diversidade genética mais comum entre alelos e podem ser usados como marcadores genéticos simples ou até mesmo para detectar as associações de um alelo com um determinado fenótipo. É importante considerar, conforme salienta Rafalski (2002), que a análise de haplótipos de SNPs fornece um meio muito mais eficiente de associar alelos com características específicas do que quando estudam-se os SNPs individualmente. Devido ao fenômeno genético-populacional do desequilíbrio de ligação, os alelos de um determinado conjunto de marcadores genéticos (haplótipos) que estão próximos a mutações causadoras de alguma variação fenotípica são também relacionados com essa variação (Kwok & Gu, 1999). É justamente a existência desta associação que abre a possibilidade de que o polimorfismo

genético responsável por um determinado fenótipo de interesse possa ser monitorado pela análise dos marcadores genéticos. Neste contexto, a descoberta de SNPs dentro ou mesmo próximos a genes de interesse representa um passo fundamental para os modernos estudos de mapeamento por associação, voltados para a identificação das causas genéticas últimas da variação fenotípica observada em caracteres de grande importância econômica.

4 MATERIAL E MÉTODOS

4.1 MINERAÇÃO DOS DADOS

A fim de se estabelecer um critério para a identificação e classificação dos genes de celulose sintase em *Eucalyptus*, a árvore filogenética de referência, publicada por Richmond & Somerville (2000), foi reconstruída, baseada nos genes da família *CesA* já conhecidos em *Arabidopsis thaliana*. As seqüências referentes a cada gene, identificadas pela literatura, foram obtidas do banco de dados *GenBank*, através do site do NCBI (http://www.ncbi.nlm.nih.gov/). Em *A. thaliana*, 10 genes da família *CesA* já foram identificados. Pelo menos um representante de cada gene foi escolhido para a confecção do dendrograma (Tabela 1). As seqüências, obtidas no formato FASTA, foram alinhadas pelo programa ClustalX (Thompson *et al.*, 1997), que realiza alinhamentos múltiplos para obtenção de árvores filogenéticas (Jeanmougin *et al.*, 1998). Para construção do alinhamento múltiplo foram utilizados os parâmetros *padrões* do programa. O dendrograma foi construído utilizando-se o algoritmo de Neighbor-Joining (Saitou & Nei, 1987). Os resultados obtidos foram visualizados através do programa TreeView (Page, 1996).

A partir do dendrograma de referência, em concordância com os resultados relatados na literatura em relação à filogenia dos genes de celulose sintase em *Arabidopsis thaliana*, três seqüências foram escolhidas: AB006703, AF088917 e AL035526, correspondentes aos genes *AtCesA4*, *AtCesA7* e *AtCesA8*, respectivamente, descritos por Holland *et al.* (2000). Essas seqüências foram utilizadas para identificar, no banco de dados de ESTs (contendo 168.741 seqüências depositadas e anotação realizada em outubro de 2005) do Projeto Genolyptus (Rede Brasileira de Pesquisa do Genoma de Eucalyptus), as seqüências homólogas a esses genes. A identificação foi realizada através da análise pelo programa BLAST (Altschul *et al.*, 1990). Um total de 320 seqüências que continham um valor-E menor que 10^{-10} foram selecionadas.

O algoritmo realizado pelo programa BLAST (*Basic Local Alignment Search Tool*) (Altschul *et al.*, 1990) é, sem dúvida, a ferramenta mais popular para a pesquisa de

seqüências em um banco de dados, executando comparações de seqüências em pares, procurando regiões de simi-laridade local (Prosdocimi *et al.*, 2002). A descoberta de seqüências similares a seqüências de DNA ou de proteínas conhecidas é a primeira pista para a identificação da função dessa seqüência. O programa fornece uma informação que é de suma importância para a interpretação dos dados: o valor-E. A grosso modo, este valor-E de um alinhamento entre pares de seqüências indica o número de alinhamentos que podem ser encontrados com uma pontuação maior ou igual à pontuação observada, em uma pesquisa feita no banco de dados com uma seqüência aleatória. Portanto, um valor-E alto (>1) indica que o alinhamento provavelmente ocorreu ao acaso e que a seqüência alvo foi alinhada com uma seqüência não relacionada do banco de dados. Valores-E menores do que 10^{-3} são comumente usados como limites em pesquisas (Gibas & Jambeck, 2001).

Tabela 1. Genes *CesA* de *Arabidopsis thaliana* utilizados como referência.

gene	nº de acesso (*GenBank*)	Referência
AtCesA1	AF027172	Holland *et al.* (2000)
AtCesA1	BT008654	NCBI
AtCesA2	AF027173	Holland *et al.* (2000)
AtCesA3	AF027174	Holland *et al.* (2000)
AtCesA3	AB018111	Scheible *et al.* (2001)
AtCesA4	AB006703	Holland *et al.* (2000)
AtCesA5	AB016893	Holland *et al.* (2000)
AtCesA6	AF062485	Holland *et al.* (2000)
AtCesA6	AB025637	Fagard *et al.* (2000)
AtCesA7	AF088917	Holland *et al.* (2000)
AtCesA7	NM_121748	NCBI
AtCesA8	AF267742	Samuga & Joshi (2002)
AtCesA8	AL035526	Holland *et al.* (2000)
AtCesA9	AC007019	Holland *et al.* (2000)
AtCesA10	AC006300	Samuga & Joshi (2002)

Um alinhamento foi realizado entre a seqüências de *A. thaliana* e os ESTs de *Eucalyptus* com o intuito de se verificar o grau de identidade existente entre tais seqüências. Para fins de controle, foram incluídas no alinhamento seqüências de genes *CesA* de outras espécies (*Pinus*, *Populus* e *Nicotiana*) que sabidamente estavam relacionadas com a produção de celulose nestas espécies (Tabela 2). A partir deste alinhamento, um dendrograma foi novamente obtido utilizando-se o algoritmo de Neighbor-Joining (Saitou & Nei, 1987). A consistência dos nós do dendrograma que deram origem aos diferentes grupos de homologia foi avaliada pelo procedimento de *bootstrap*, utilizando 1000 reamostragens. O procedimento conhecido como *bootstrap* é um método de reamostragem que se constitui, nesse caso, na reconstrução de um grande número de dendrogramas utilizando seqüências obtidas por reamostragem, com reposição, dos sítios do alinhamento original. Este procedimento permite a avaliação da consistência dos nós de um dendrograma frente às variações presentes ao longo da seqüência analisada. O grau de consistência dos nós é dado pela proporção em que esses nós ocorrem nos dendrogramas obtidos em cada réplica do processo de reamostragem (Felsenstein, 2004).

Tabela 2. Genes *CesA* de outras espécies utilizados como referência.

gene	nº de acesso (*GenBank*)	referência	espécie
CesA7-like	AY330166.1	NCBI	*Populus tremuloides*
CesA5-like	AY330165	NCBI	*Populus tremuloides*
CesA	BX249523	NCBI	*Pinus pinaster*
CesA	BX682714	NCBI	*Pinus pinaster*
CesA	AL750979	NCBI	*Pinus pinaster*
CesA4	AY162181	Kalluri & Joshi (2004)	*Populus tremuloides*
CesA	AF233892	NCBI	*Nicotiana tabacum*
CesA2	AF233893	NCBI	*Nicotiana tabacum*
CesA2	AY573572	NCBI	*Populus tremula x Populus tremuloides*
CesA7	AY162180	Samuga & Joshi (2004)	*Populus tremuloides*
CesA4	AY573575	*NCBI*	*Populus tremula x Populus tremuloides*
CesA3	AY573573	*NCBI*	*Populus tremula x Populus tremuloides*
CesA1	AY573571	*NCBI*	*Populus tremula x Populus tremuloides*
CesA3	AF527387	Kalluri & Joshi (2004)	*Populus tremuloides*
CesA6	AY196991	Samuga & Joshi (2004)	*Populus tremuloides*
CesA	BX784298	NCBI	*Pinus pinaster*
CesA1	AY639654	NCBI	*Pinus radata*
CesA	BX255542	NCBI	*Pinus pinaster*
CesA	BX254358	NCBI	*Pinus pinaster*
CesA	BX252761	NCBI	*Pinus pinaster*
CesA	BX250396	NCBI	*Pinus pinaster*
CesA2	AY095297	Samuga & Joshi (2004)	*Populus tremuloides*

Com base no alinhamento realizado com as seqüências de cada um desses grupos, isoladamente, foi possível detectar genes, em *Eucalyptus*, que são homólogos às isoformas dos genes *CesA* de *Arabidopsis thaliana* relacionados à produção de celulose da parede celular secundária (*AtCesA4*, *AtCesA7* e *AtCesA8*).

4.2 DESENHO DE PRIMERS

Para fins de desenho de *primers* específicos, destinados a amplificação e/ou seqüenciamento dos diferentes genes de celulose sintase identificados em *Eucalyptus*, foram utilizados os alinhamentos obtidos para cada um dos grupos de homologia relacionados à síntese de celulose em parede secundária definidos em *Arabidopsis thaliana* (*AtCesA4*, *AtCesA7* e *AtCesA8*).

A fim de se minimizar os problemas com a utilização dos *primers* para amplificação de fragmentos genômicos dos genes de celulose sintase em *Eucalyptus*, foi realizada uma análise buscando se detectar e posicionar os íntrons existentes nestes genes. Esta análise consistiu na comparação da seqüência consenso de cada alinhamento com a versão genômica correspondente aos genes homólogos de *Arabidopsis thaliana*, através da utilização do programa EST2GENOME (Mott, 1997), disponível em http://bioweb.pasteur.fr/-seqanal/interfaces/est2genome.html. Com este programa é possível visualizar a posição estimada do íntron existente na seqüência genômica.

Um total de 28 *primers* foram desenhados com o auxílio do programa Primer3 (Rozen & Skaletsky, 2000), disponível em http://frodo.wi.mit.edu/cgibin/primer3/-primer3_www.cgi. Foram utilizados os valores *padrões* do programa. Estes *primers* foram desenhados com o intuito de se amplificar e/ou seqüenciar regiões específicas dos genes de celulose sintase preferencialmente expressos na parede secundária em espécies de *Eucalyptus*, de modo a viabilizar os estudos destinados a obter a seqüência completa destes genes ou para fins de avaliação do polimorfismo existente nestes genes entre diferentes espécies do gênero.

4.3 TRIAGEM DE UMA BIBLIOTECA DE BACs EM BUSCA DE CLONES CONTENDO GENES HOMÓLOGOS A *AtCesA7*

Dentre os *primers* desenhados para os genes homólogos a *AtCesA7*, aqueles que flanqueavam regiões internas do gene foram reunidos aos pares para a realização de uma triagem em uma biblioteca de BACs, construída no âmbito do projeto Genolyptus, em busca de clones contendo genes homólogos a *AtCesA7*.

A biblioteca de BACs utilizada nesse trabalho para identificação dos clones de interesse foi construída no laboratório do Dr. Sérgio Hermínio Brommonschenkel, na Universidade Federal de Viçosa. Esta biblioteca foi desenvolvida a partir de um indivíduo de *Eucalyptus grandis* (clone G44). A biblioteca, denominada de EUGR-BC-002, é composta por 20160 clones (armazenados em 210 placas de 96 poços), os quais possuem tamanhos médios de insertos entre 120 a 150 kb. Estima-se que esta biblioteca ofereça uma cobertura de quatro vezes do genoma (Santos, 2005; Vaz, 2005). Para fins de triagem, essa biblioteca foi organizada na forma hierarquizada de *pools* e *superpools*. Cada *pool* foi formado por 96 clones de uma placa de BACs que foram purificados juntos. Os *superpools* foram formados por conjuntos de 10 *pools* cada, perfazendo um total de 21 *superpools*. Essa estratégia de hierarquização de BACs (figura 05) teve como objetivo facilitar a detecção do clone alvo, reduzindo os gastos com a realização das PCRs (Vaz, 2005).

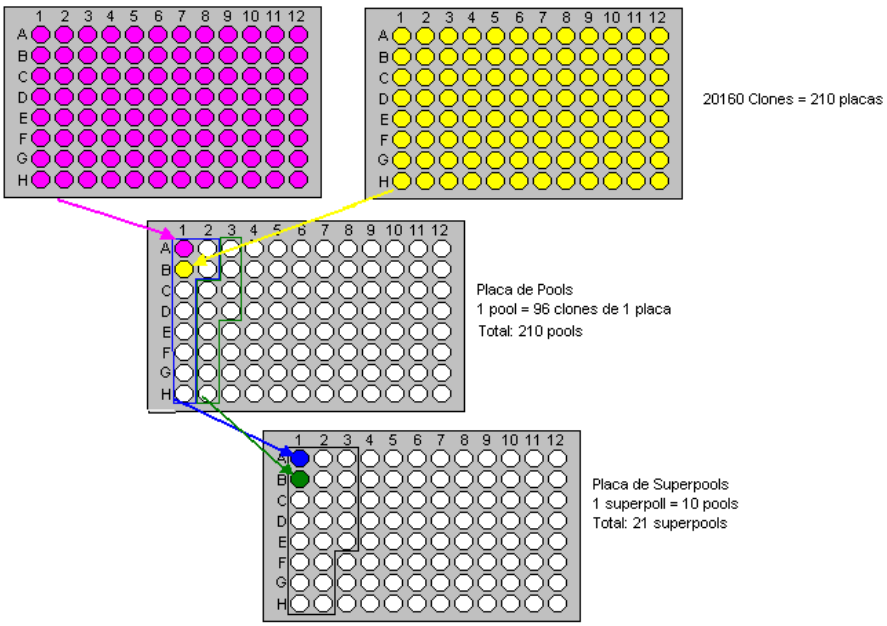


Figura 5. Estrutura em hierarquia de placas, *pools* e *superpools* da biblioteca de E*ucalyptus grandis.*

O conjunto de *pools* e *superpools* obtidos com a biblioteca original de BACs, desenvolvido por Vaz (2005), disponível no Laboratório de Biotecnologia da Embrapa Arroz e Feijão, serviu de base para a triagem realizada com os pares de *primers* específicos do gene de celulose sintase de *Eucalyptus* homólogo ao gene *AtCesA7*, de *Arabidopsis thaliana*. A triagem foi inicialmente realizada nos 21 *superpools*, através de PCR. Nesta etapa da triagem, 22 reações foram feitas para cada par de *primer*: 21 utilizando o DNA dos *superpools* e uma reação referente a um DNA controle (clone G44 de *Eucalyptus grandis*). As reações foram realizadas com um volume final de 15μL, em tampão contendo KCl (50mM), Tris-HCl (10mM, pH 8,4), Triton X-100 (0,1%), na presença de BSA (0,2μg/μL); $MgCl_2$ (1,33mM), dNTPs (0,16 μM); 2,50 pmoles de cada *primer* (*forward* e *reverse*); 0,6U da enzima Taq DNA polimerase (Phoneutria) e 6 ng de DNA molde. As reações de amplificação foram realizadas em um termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.) seguindo os seguintes passos: desnaturação a 94ºC durante 5 minutos; 30 ciclos compostos de um passo de 94ºC por 1 minuto, temperatura de anelamento dos *primers* (56ºC) durante 1 minuto, extensão a 72ºC por 1 minuto; seguidos de um passo final de extensão a 72ºC por 30 minutos. Após a amplificação, a triagem dos *superpools* foi realizada pela visualização dos produtos obtidos com eletroforese em gel de agarose (1,5%) corado com brometo de etídeo (0,2 mg/mL). A eletroforese foi realizada em cuba horizontal com voltagem constante de 120V por aproximadamente 1 hora e 30 minutos, com gel totalmente submerso em TBE 1X (Tris-HCl, 500 mM; ácido bórico, 60 mM; EDTA, 83 mM). Os resultados obtidos foram registrados utilizando-se o fotodocumentador EagleEye (Strategene), utilizando luz ultravioleta para visualizar os produtos corados com brometo de etídeo.

Após a identificação dos *superpools* positivos, a triagem foi feita dentro de cada um dos 10 *pools* correspondentes. Em uma etapa final, após a identificação dos *pools* positivos, a triagem foi então realizada em cada placa correspondente, para a identificação do clone BAC específico, contendo o gene de interesse.

4.4 ISOLAMENTO DOS BACs CONTENDO O GENE DE CELULOSE SINTASE DE *Eucalyptus* HOMÓLOGO A *AtCesA7*

A fim de se proceder o isolamento dos BACs contendo os genes de interesse, um pré-inóculo dos clones positivos foi inicialmente realizado. Os clones positivos foram repicados a partir das placas-estoque, armazenadas à -80ºC, para placas contendo meio LB sólido, na presença do antibiótico Cloranfenicol (12,5mg/mL). As placas foram incubadas em estufa a 37ºC *overnight*. Em seguida, os clones foram cultivados em 5mL de meio LB líquido, também na presença de Cloranfenicol, a 37ºC por 20h sob agitação constante de 250rpm.

Os tubos foram então centrifugados a 5500rpm por 5 minutos à temperatura ambiente para que as células fossem sedimentadas. O sobrenadante foi descartado e as células ressuspendidas em 600µL de solução GET (Glicose, 1%; EDTA, 0,01M, pH8,0; Tris-HCl, 0,025M, pH7,4) acrescida de RNAse.

Após homogeinização e incubadação a 0ºC por 5 minutos, aos tubos foram adicionados 1,2 mL de solução de lise (NaOH, 0,02M; SDS 1%). O conteúdo dos tubos foi então novamente homogeneizado com cuidado e incubado a 0ºC por mais 5 minutos. Posteriormente, 900 µL da solução de neutralização (KOAc, 3M; ácido acético glacial, 11,5%) foram adicionados, com nova homogeinização (com cuidado), seguida de incubação a 0ºC por mais 15 minutos.

Transcorrida a neutralização, os tubos foram centrifugados a 5000 rpm por 15 minutos a 4ºC. Um volume de 750 µL do sobrenadante foi transferido para três tubos de 1,5mL e a estes foram adicionados 450 µL de isopropanol a -20ºC para precipitação do DNA. Os tubos foram homogeneizados em *vortex* e posteriormente centrifugados a 14000rpm durante 5 minutos. O sobrenadante foi descartado e as amostras foram lavadas com 500µL de etanol 70%, centrifugando os tubos a 14000rpm por 2 minutos e descartando-se o álcool. Depois de obter as amostras secas (*pellets*), as mesmas foram dissolvidas em 30 µL de H_2O e estocadas em *freezer* a -20ºC.

4.5 SEQÜENCIAMENTO DO GENE DE CELULOSE SINTASE DE *Eucalyptus* HOMÓLOGO A *AtCesA7*

Os BACs, isolados, contendo os genes de celulose sintase de *Eucalyptus* homólogos ao gene *AtCesA7* de *Arabidopsis thaliana*, foram submetidos à PCR para amplifica-

ção de fragmentos do gene. As reações foram realizadas com um volume final de 12μL, em tampão contendo KCl (50mM), Tris-HCl (10mM, pH 8,4), Triton X-100 (0,1%), na presença de BSA (0,25μg/μL); $MgCl_2$ (1,66mM), dNTPs (0,25 μM); 0,50 pmol de cada *primer* (*forward* e *reverse*); 1,5U da enzima Taq DNA polimerase (Phoneutria) e 6 ng de DNA molde.

As reações de amplificação foram realizadas em um termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.) seguindo os seguintes passos: desnaturação a 94ºC durante 5 minutos; 45 ciclos compostos de um passo de 94ºC por 1 minuto, temperatura de anelamento dos *primers* (56ºC) durante 1 minuto, extensão a 72ºC por 1 minuto; seguidos de um passo final de extensão a 72ºC por 30 minutos.

O protocolo utilizado torna a etapa de purificação do produto de PCR desnecessária, pois utiliza uma menor quantidade de *primers* na reação e um maior número de ciclos para amplificação. Espera-se, neste caso, que todo o *primer* presente na reação seja consumido, de modo a não prejudicar o desenvolvimento das reações de seqüenciamento subseqüentes, o que geraria resultados insatisfatórios.

As amplificações dos fragmentos foram confirmadas em gel de agarose 1,5% corado com brometo de etídeo e visualizadas em fotodocumentador através de luz ultravioleta.

Os fragmentos amplificados com os pares de *primers* foram seqüenciados com cada um dos *primers* separadamente (*forward* e *reverse*). As reações de seqüenciamento seguiram o seguinte protocolo: 1μL do produto de PCR, 2,5 picomoles do *primer forward* ou *reverse*, 2μL de tampão *Save-Money* (Tris-HCl, 200mM, pH9,0; $MgCl_2$, 0,5mM), 1μL do kit DYEnamic^{TM} ET Terminator Cycle Sequencing (Pharmacia Biotech, EUA), para um volume final de 10μL. As reações de seqüenciamento foram realizadas em termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.), seguindo o seguinte protocolo: 30 ciclos de 95º por 20 segundos e 60ºC por 1minuto e 15 segundos.

O produto da reação de seqüenciamento foi então submetido à etapa de precipitação da seguinte maneira: 40μL de isopropanol 65% foram adicionados aos 10μL da reação. A mistura foi homogeneizada em *vortex* por 30 segundos. Após um período de incubação à temperatura ambiente por 20 minutos, as amostras foram centrifugadas a 2000rcf por 45 minutos e o sobrenadante foi descartado. Ao precipitado, foram adicionados 250μL de etanol 60%, para a lavagem do material, seguindo-se uma centrifugação a 2000rcf por 10 minutos. O processo de lavagem com etanol foi repetido utilizando 100μL do mesmo, e o sobrenadante, novamente descartado. A placa contendo as amostras foi então centrifuga-

da de maneira invertida a 500rpm por 1 minuto para garantir a retirada de todo o etanol. Uma secagem complementar foi realizada deixando-se a placa por aproximadamente 10 minutos em capela de fluxo laminar.

O *pellet* foi ressuspendido em 10μL de formamida (Hi-Di) e depois submetido a um processo de desnaturação em termociclador a 95ºC por 5 minutos, seguido de imediata incubação em gelo por aproximadamente 3 minutos. As amostras foram então submetidas à análise pelo equipamento ABI-3100 (Applied Biosystems, Inc.).

Durante a etapa de desenho dos *primers* foram desenhados oligos orientados para as extremidades do alinhamento, a fim de se seqüenciar uma maior extensão do gene, tanto no sentido 5' quanto no 3'. O seqüenciamento nestes casos foi realizado utilizando o próprio BAC como molde. Estas reações de seqüenciamento seguiram o seguinte protocolo: 15,8μL da solução obtida no isolamento do BAC, 10,24 picomoles do *primer forward* ou *reverse*, 10μL de tampão *Save-Money* (Tris-HCl, 200mM, pH9,0; $MgCl_2$, 0,5mM), 2μL do kit Big Dye™ Terminator v3.0 Ready Reaction Cycle Sequencing (Applied Biosystems, Inc.), para um volume final de 25μL. As reações de seqüenciamento foram realizadas em termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.), seguindo o seguinte protocolo: 95ºC por cinco minutos, seguidos de 95 ciclos de 95ºC por 30 segundos, 56ºC por dez segundos e 60ºC por dois minutos e trinta segundos. As etapas de precipitação, desnaturação e análise no equipamento ABI-3100 seguiram os mesmos passos já descritos para o seqüenciamento dos produtos de PCR.

4.6 ANÁLISE DAS SEQÜÊNCIAS

Os fragmentos seqüenciados foram submetidos à análise de *base-calling* pelo programa PHRED (Ewing *et al.*, 1998). Este programa interpreta a informação representada nos eletroferogramas em termos da seqüência de bases e associa a cada base uma medida de qualidade (Ewing *et al.*, 1998).

As seqüências obtidas pela utilização dos diferentes *primers* nos diferentes BACs identificados na etapa de *triagem* foram analisadas utilizando-se o programa *Staden Package* (Staden *et al.*, 2000).

4.7 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DO GENE DE CELULOSE SINTASE DE *Eucalyptus* ORTÓLOGO A *AtCesA7*

A abordagem utilizada para se avaliar preliminarmente os níveis de expressão do gene de celulose sintase de *Eucalyptus* homólogo ao gene *AtCesA7* de *Arabidopsis thaliana*, em diferentes condições consistiu na comparação das freqüências de ESTs deste gene nas diferentes bibliotecas de cDNA integrantes do banco de dados de ESTs do projeto Genolyptus. A significância estatística das diferenças entre as freqüências observadas de ESTs nas diferentes bibliotecas foi avaliada conforme preconizado por Audic & Claverie (1997). Foram comparadas as freqüências de ESTs do gene de interesse nas bibliotecas de cDNA construídas a partir de RNA obtido de xilema, folhas maduras e plântulas. Foram utilizados os procedimentos baseados em randomização para obter os p-valores associados a cada comparação. Foram comparados os níveis de expressão do gene de interesse em xilema de *Eucalyptus grandis* e *Eucalyptus globulus*, e ainda os níveis de expressão em xilema e folhas de *Eucalyptus grandis*.

4.8 DETECÇÃO DE POLIMORFISMO NO GENE *CesA7.2* DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO *Eucalyptus*

Para fins de avaliação do polimorfismo presente no gene *CesA7.2* de celulose sintase de *Eucalyptus* homólogo ao gene *AtCesA7* foram seqüenciados fragmentos deste gene em 6 indivíduos representantes de 4 diferentes espécies de *Eucalyptus* (*E. camadulensis, E. dunni, E. urophylla, E. globulus*) e ainda de 1 híbrido interespecífico (*E. dunni x E. grandis*).

A extração do DNA destes indivíduos foi realizada a partir das folhas previamente estocadas em freezer a -80ºC. A extração cumpriu o protocolo descrito por Ferreira & Grattapaglia (1998). Cerca de 150 mg de tecido foliar foram colocados em microtubos de 1,5mL para adição de nitrogênio líquido. Ao conteúdo macerado foram adicionados 700µL de tampão de extração CTAB (CTAB, 2%; EDTA, 20mM; Tris-HCl, 100mM, pH8,0). Os microtubos foram incubados em banho-maria a 65ºC por aproximadamente 30 minutos, sendo os tubos agitados a cada 10 minutos. Uma vez resfriados à temperatura ambiente, foram então adicionados aos tubos 600µL de CIA (clorofórmio-álcool isoamíli-

co, 24:1). Após agitação cuidadosa e completa homogeinização, os tubos foram centrifugados a 1500rpm por 5 minutos. A fase aquosa superior foi retirada e transferida para um novo tubo e a esta foram adicionados 50 μL de solução contendo CTAB (10%) e NaCl (1,4M). Uma vez homogeinizados, procedeu-se então nova extração com 600 μL de CIA. A fase aquosa superior foi então transferida para um novo tubo e a esta foram acrescentados 400 μL de isopropanol frio. Esta mistura foi estocada a -20ºC por aproximadamente 12 horas. Após esse período os tubos foram centrifugados a 7500 rpm por 5 minutos; o sobrenadante foi descartado gentilmente. O *pellet* formado foi lavado 2 vezes com etanol 70% durante 5 minutos e uma vez com etanol absoluto por 3 minutos. Os *pellets* devidamente secos foram ressuspendidos em 50 μL de tampão TE (Tris-HCl, 10mM, pH8,0; EDTA, 1mM).

A quantificação do DNA total obtido na extração foi realizada em gel de agarose a 1,5% corado com brometo de etídeo. Um total de 10μL da solução estoque obtida na extração do DNA foi aplicado no gel, conjuntamente com quantidade conhecidas de DNA genômico do fago λ (Invitrogen). A eletroforese foi realizada em cuba horizontal sob voltagem constante de 100V. Os resultados obtidos foram digitalizados utilizando-se um fotodocumentador KODAK EDAS 290, com a utilização de luz ultravioleta. Após a quantificação do DNA, as amostras foram diluídas de modo a fazer com que a concentração final de todas as amostras atingisse aproximadamente 2,5ng/μL.

O DNA genômico dos indivíduos foi então submetido à PCR utilizando pares de *primers* desenhados especificamente para o grupo de homologia com o gene *AtCesA7*. As reações foram realizadas com um volume final de 12μL, em tampão contendo KCl (50mM), Tris-HCl (10mM, pH 8,4), Triton X-100 (0,1%), na presença de BSA (0,25μg/μL); $MgCl_2$ (1,66mM), dNTPs (0,25 μM); 0,50 pmol de cada *primer* (*forward* e *reverse*); 1,5U da enzima Taq DNA polimerase (Phoneutria) e 6 ng de DNA molde. As reações de amplificação foram realizadas em um termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.) seguindo os seguintes passos: desnaturação a 94ºC durante 5 minutos; 45 ciclos compostos de um passo de 94ºC por 1 minuto, temperatura de anelamento dos *primers* (56ºC) durante 1 minuto, extensão a 72ºC por 1 minuto; seguidos de um passo final de extensão a 72ºC por 30 minutos. As amplificações dos fragmentos foram confirmadas em gel de agarose 1,5% corado com brometo de etídeo e visualizadas em fotodocumentador através de luz ultravioleta.

Após a confirmação da amplificação, os fragmentos foram então seqüenciados a partir de cada um dos *primers* separadamente (*forward* e *reverse*). As reações de seqüen-

ciamento seguiram o seguinte protocolo: 1μL do produto de PCR, 2,5 picomoles do *primer forward* ou *reverse*, 2μL de tampão *Save-Money* (Tris-HCl, 200mM, pH9,0; $MgCl_2$, 0,5mM), 1μL do kit DYEnamic^{TM} ET Terminator Cycle Sequencing (Pharmacia Biotech, EUA), para um volume final de 10μL. As reações de seqüenciamento foram realizadas em termociclador GeneAmp PCR System 9700 (Applied Biosystems, Inc.), seguindo o seguinte protocolo: 30 ciclos de 95º por 20 segundos e 60ºC por 1minuto e 15 segundos).

O produto da reação de seqüenciamento foi então submetido à etapa de precipitação da seguinte maneira: 40μL de isopropanol 65% foram adicionados aos 10μL da reação. A mistura foi homogeneizada em *vortex* por 30 segundos. Após um período de incubação à temperatura ambiente por 20 minutos, as amostras foram centrifugadas a 2000rcf por 45 minutos e o sobrenadante foi descartado. Ao precipitado, foram adicionados 250μL de etanol 60%, para a lavagem do material, seguindo-se uma centrifugação a 2000rcf por 10 minutos. O processo de lavagem com etanol foi repetido utilizando 100μL do mesmo, e o sobrenadante, novamente descartado. A placa contendo as amostras foi então centrifugada de maneira invertida a 500rpm por 1 minuto para garantir a retirada de todo o etanol. Uma secagem complementar foi realizada deixando-se a placa por aproximadamente 10 minutos em capela de fluxo laminar.

O *pellet* foi ressuspendido em 10μL de formamida (Hi-Di) e depois submetido a um processo de desnaturação em termociclador a 95ºC por 5 minutos, seguido de imediata incubação em gelo por aproximadamente 3 minutos. As amostras foram então submetidas à análise pelo equipamento ABI-3100 (Applied Biosystems, Inc.).

Os fragmentos seqüenciados foram então submetidos à análise de *base-calling* pelo programa PHRED (Ewing *et al.*, 1998). As seqüências obtidas e os valores de qualidade a elas associados foram analisados pela utilização do programa Staden Package (Staden *et al.*, 2000) para detecção dos prováveis SNPs, conforme metodologia descrita em Bonfield *et al*. (1998). Os sítios polimórficos foram caracterizados em termos do tipo de mutação detectada e de região de ocorrência (íntrons ou éxons). Foram ainda estimados os níveis de diversidade nucleotídica (Nei, 1987), utilizando-se o software Mega versão 3.1 (Kumar *et al*., 2004).

5 RESULTADOS E DISCUSSÃO

5.1 MINERAÇÃO DOS DADOS

A Figura 6 apresenta a árvore filogenética de referência obtida com as seqüências de *Arabidopsis thaliana*, reproduzindo os dados da literatura em relação à filogenia dos genes de celulose sintase desta espécie modelo. O dendrograma foi gerado pelo algoritmo de Neighbor-Joining (Saitou & Nei, 1987), tendo sido a consistência dos nós avaliada pela realização de 1000 ciclos de *bootstrap*. Os genes *AtCesA4*, *AtCesA7* e *AtCesA8*, representados neste dendrograma, são responsáveis pela formação de celulose na parede celular secundária em *A. thaliana* (Joshi *et al*., 2004).

O dendrograma produzido incluindo-se as seqüências obtidas com base nos resultados de busca por genes similares aos genes *AtCesA4, AtCesA7* e *AtCesA8* no banco de dados de ESTs de *Eucalyptus* do projeto Genolyptus, além de seqüências de genes de celulose sintase de outras espécies, utilizadas também como referência (Tabela 2, página 20), está apresentado na Figura 7. Inicialmente, observa-se que para os genes de referência houve concordância com os achados na literatura. Segundo os resultados de Samuga & Joshi (2004), por exemplo, o gene *PtrCesA1* é ortólogo ao gene *AtCesA8*, enquanto o gene *PtrCesA2* apresenta ortologia ao gene *AtCesA7*. Concordâncias desta natureza sugerem que a metodologia utilizada se mostrou adequada para classificar os ESTs de *Eucalyptus* em grupos de ortologia dentro da família de celulose sintase. A lista dos ESTs de *Eucalyptus* identificados como relacionados a genes de celulose sintase de parede secundária está apresentada na Tabela 3. Observa-se ainda que, com base nos resultados obtidos por este dendrograma, os genes ortólogos se revelaram mais similares do que os genes parálogos, indicando a ausência de mecanismos evolutivos conhecidos como evolução em concerto, em que cópias parálogas tendem a se homogeinizar, mantendo um baixo grau de divergência entre si (Brown, 2002). Pelo grau de divergência filogenética apresentado entre as espécies representadas no dendrograma, é possível ainda inferir que o papel que os genes de celulose sintase realizam atualmente deve ter sido estabelecido há muito tempo no proces-

so evolutivo de plantas superiores e que esses genes, mesmo estando relacionados e interagindo durante a produção da parede secundária, sofreram uma intensa especificação funcional durante o processo evolutivo, conforme salientam Doblin *et al*. (2002).

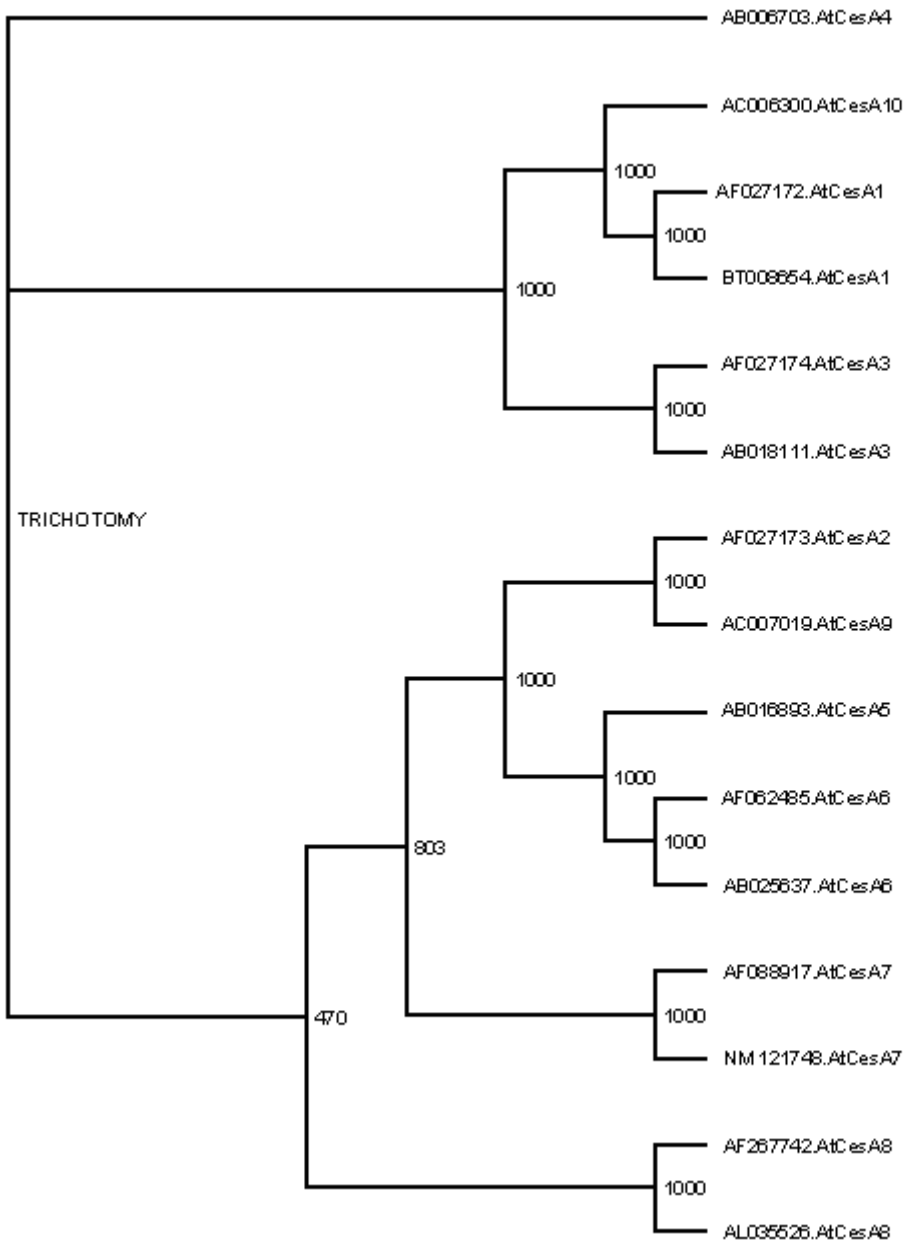
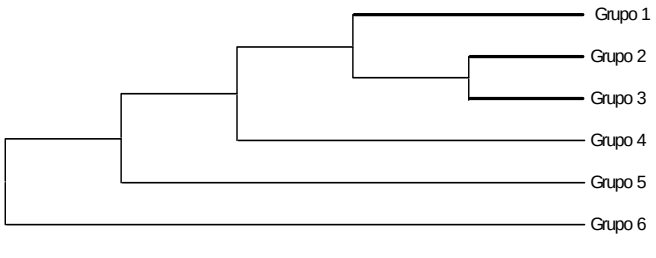
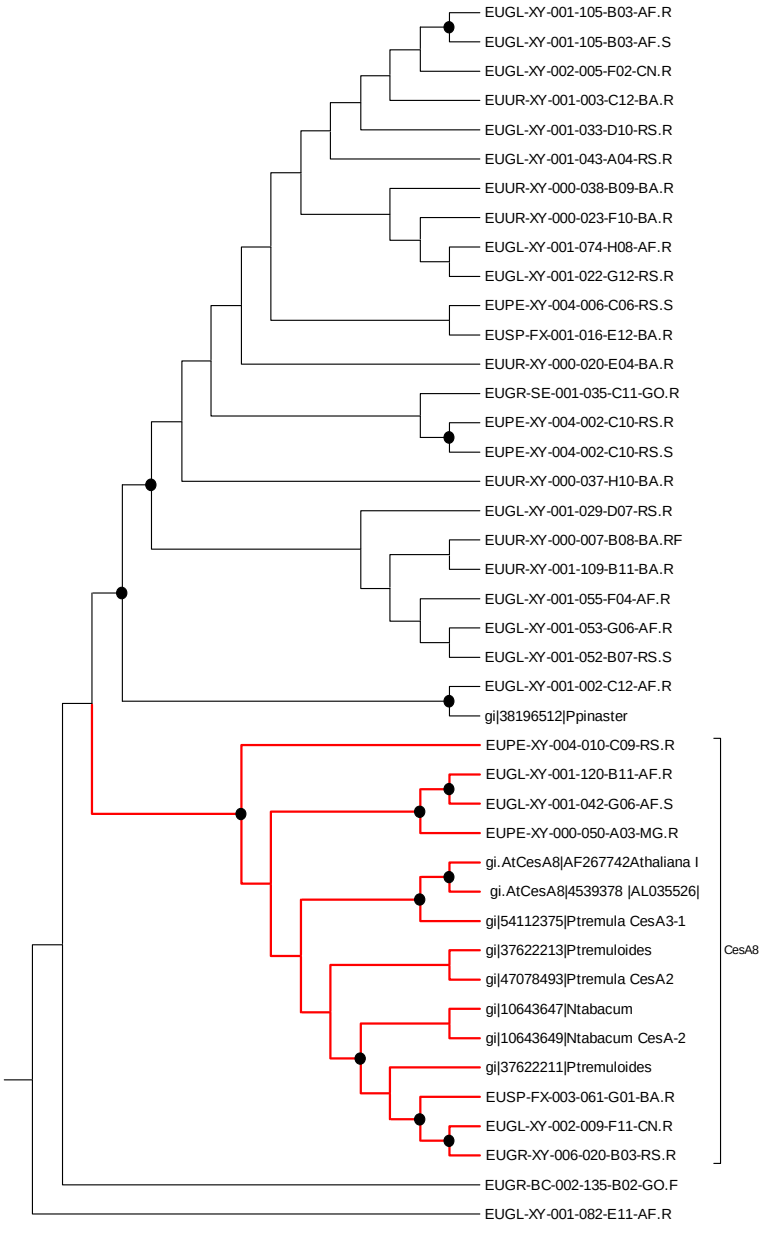


Figura 6. Dendrograma de referência construído utilizando o algorítmo de Neighbor-Joining representando as relações entre genes de celulose sintase de *Arabidopsis thaliana* (Tabela 1).

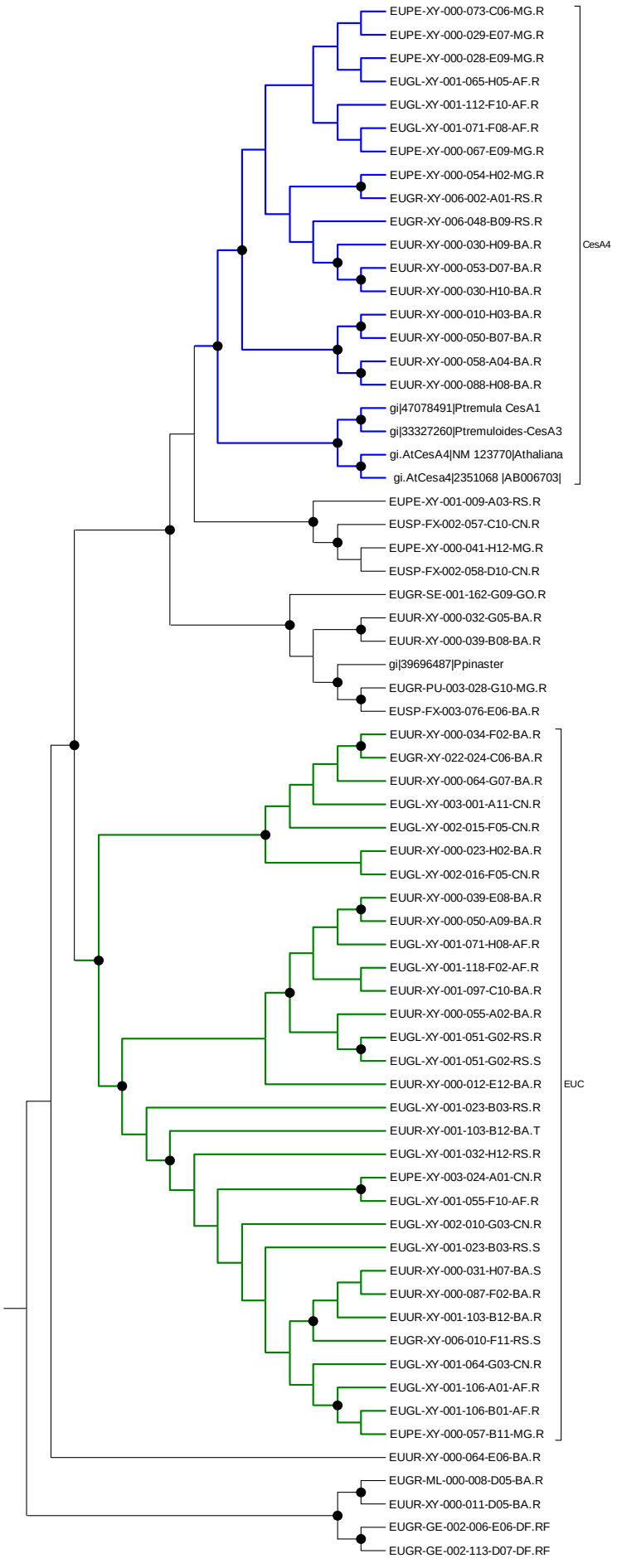


(A)

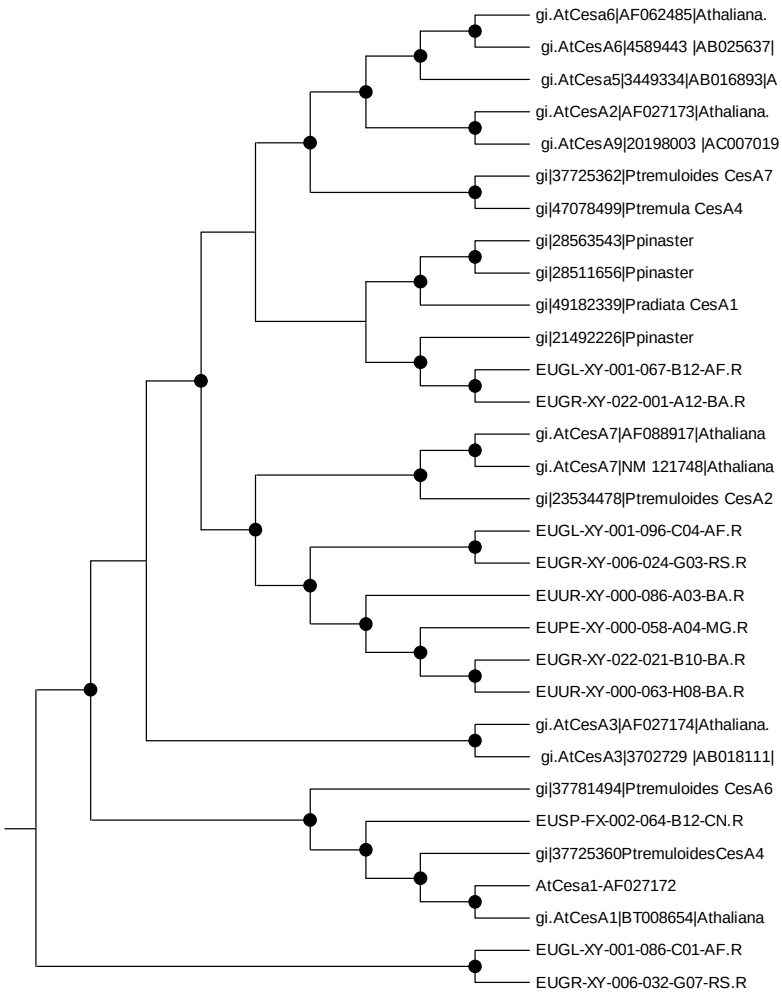


(B)

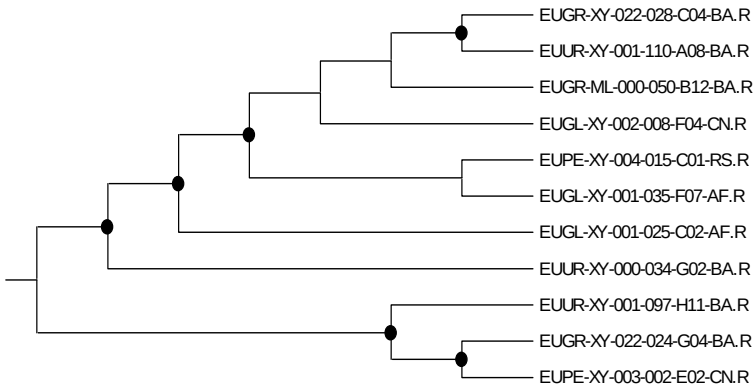
Figura 07 continua...


EUPE-XY-000-073-C06-MG.R
EUPE-XY-000-029-E07-MG.R
EUPE-XY-000-028-E09-MG.R
EUGL-XY-001-065-H05-AF.R
EUGL-XY-001-112-F10-AF.R
EUGL-XY-001-071-F08-AF.R
EUPE-XY-000-067-E09-MG.R
EUPE-XY-000-054-H02-MG.R
EUGR-XY-006-002-A01-RS.R
EUGR-XY-006-048-B09-RS.R
EUUR-XY-000-030-H09-BA.R
EUUR-XY-000-053-D07-BA.R
EUUR-XY-000-030-H10-BA.R
EUUR-XY-000-010-H03-BA.R
EUUR-XY-000-050-B07-BA.R
EUUR-XY-000-058-A04-BA.R
EUUR-XY-000-088-H08-BA.R
gi|47078491|Ptremula CesA1
gi|33327260|Ptremuloides-CesA3
gi.AtCesA4|NM 123770|Athaliana
gi.AtCesa4|2351068 |AB006703|
CesA4
EUPE-XY-001-009-A03-RS.R
EUSP-FX-002-057-C10-CN.R
EUPE-XY-000-041-H12-MG.R
EUSP-FX-002-058-D10-CN.R
EUGR-SE-001-162-G09-GO.R
EUUR-XY-000-032-G05-BA.R
EUUR-XY-000-039-B08-BA.R
gi|39696487|Ppinaster
EUGR-PU-003-028-G10-MG.R
EUSP-FX-003-076-E06-BA.R
EUUR-XY-000-034-F02-BA.R
EUGR-XY-022-024-C06-BA.R
EUUR-XY-000-064-G07-BA.R
EUGL-XY-003-001-A11-CN.R
EUGL-XY-002-015-F05-CN.R
EUUR-XY-000-023-H02-BA.R
EUGL-XY-002-016-F05-CN.R
EUUR-XY-000-039-E08-BA.R
EUUR-XY-000-050-A09-BA.R
EUGL-XY-001-071-H08-AF.R
EUGL-XY-001-118-F02-AF.R
EUUR-XY-001-097-C10-BA.R
EUUR-XY-000-055-A02-BA.R
EUGL-XY-001-051-G02-RS.R
EUGL-XY-001-051-G02-RS.S
EUUR-XY-000-012-E12-BA.R
EUGL-XY-001-023-B03-RS.R
EUUR-XY-001-103-B12-BA.T
EUGL-XY-001-032-H12-RS.R
EUPE-XY-003-024-A01-CN.R
EUGL-XY-001-055-F10-AF.R
EUGL-XY-002-010-G03-CN.R
EUGL-XY-001-023-B03-RS.S
EUUR-XY-000-031-H07-BA.S
EUUR-XY-000-087-F02-BA.R
EUUR-XY-001-103-B12-BA.R
EUGR-XY-006-010-F11-RS.S
EUGL-XY-001-064-G03-CN.R
EUGL-XY-001-106-A01-AF.R
EUGL-XY-001-106-B01-AF.R
EUPE-XY-000-057-B11-MG.R
EUC
EUUR-XY-000-064-E06-BA.R
EUGR-ML-000-008-D05-BA.R
EUUR-XY-000-011-D05-BA.R
EUGR-GE-002-006-E06-DF.RF
EUGR-GE-002-113-D07-DF.RF

(C)

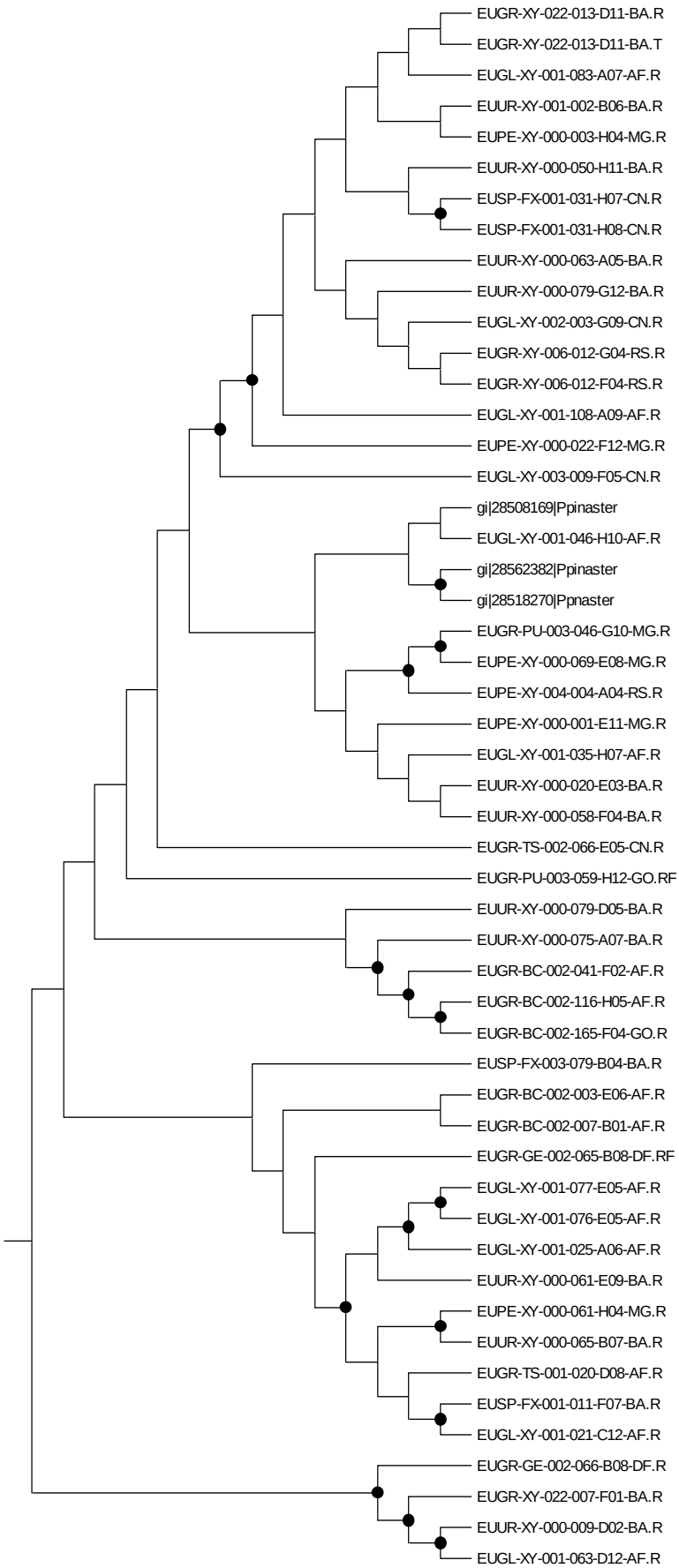


(D)



(E)

Figura 07 continua...


EUGR-XY-022-013-D11-BA.R
EUGR-XY-022-013-D11-BA.T
EUGL-XY-001-083-A07-AF.R
EUUR-XY-001-002-B06-BA.R
EUPE-XY-000-003-H04-MG.R
EUUR-XY-000-050-H11-BA.R
EUSP-FX-001-031-H07-CN.R
EUSP-FX-001-031-H08-CN.R
EUUR-XY-000-063-A05-BA.R
EUUR-XY-000-079-G12-BA.R
EUGL-XY-002-003-G09-CN.R
EUGR-XY-006-012-G04-RS.R
EUGR-XY-006-012-F04-RS.R
EUGL-XY-001-108-A09-AF.R
EUPE-XY-000-022-F12-MG.R
EUGL-XY-003-009-F05-CN.R
gi|28508169|Ppinaster
EUGL-XY-001-046-H10-AF.R
gi|28562382|Ppinaster
gi|28518270|Ppnaster
EUGR-PU-003-046-G10-MG.R
EUPE-XY-000-069-E08-MG.R
EUPE-XY-004-004-A04-RS.R
EUPE-XY-000-001-E11-MG.R
EUGL-XY-001-035-H07-AF.R
EUUR-XY-000-020-E03-BA.R
EUUR-XY-000-058-F04-BA.R
EUGR-TS-002-066-E05-CN.R
EUGR-PU-003-059-H12-GO.RF
EUUR-XY-000-079-D05-BA.R
EUUR-XY-000-075-A07-BA.R
EUGR-BC-002-041-F02-AF.R
EUGR-BC-002-116-H05-AF.R
EUGR-BC-002-165-F04-GO.R
EUSP-FX-003-079-B04-BA.R
EUGR-BC-002-003-E06-AF.R
EUGR-BC-002-007-B01-AF.R
EUGR-GE-002-065-B08-DF.RF
EUGL-XY-001-077-E05-AF.R
EUGL-XY-001-076-E05-AF.R
EUGL-XY-001-025-A06-AF.R
EUUR-XY-000-061-E09-BA.R
EUPE-XY-000-061-H04-MG.R
EUUR-XY-000-065-B07-BA.R
EUGR-TS-001-020-D08-AF.R
EUSP-FX-001-011-F07-BA.R
EUGL-XY-001-021-C12-AF.R
EUGR-GE-002-066-B08-DF.R
EUGR-XY-022-007-F01-BA.R
EUUR-XY-000-009-D02-BA.R
EUGL-XY-001-063-D12-AF.R

(F)

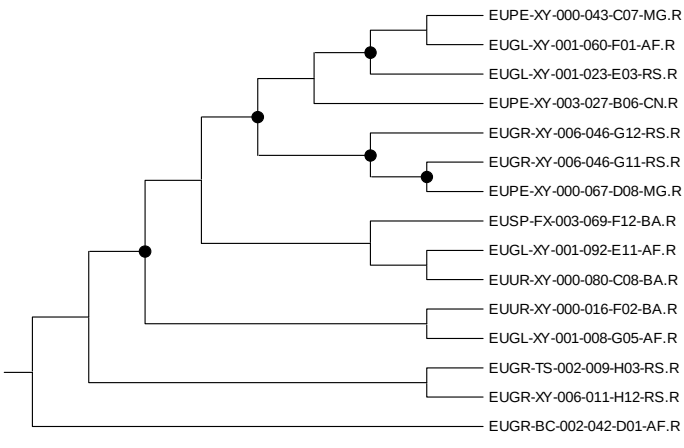


Figura 7. Dendrogramas representativos das relações entre ESTs de *Eucalyptus* e seqüências de celulose sintase de outras espécies, utilizadas como referência (Tabelas 1 e 2), dendrograma resumido (A); grupo 1 (B); grupo 2 (C); grupo 3 (D); grupo 4 (E); grupo 5 (F) e grupo 6 (G).

Dos dendrogramas exibidos na Figura 7, quatro grupos de ortologia foram selecionados. Os três primeiros grupos foram formados pelos genes *AtCesA4*, *AtCesA7* e *AtCesA8*, responsáveis pela produção de celulose sintase na parede celular secundária, e os ESTs de *Eucalyptus* a eles relacionados. O quarto grupo foi constituído exclusivamente de ESTs de *Eucalyptus*, mas foi selecionado por ter se apresentado próximo aos genes dos outros três grupos.

Tabela 3. Relação de ESTs de *Eucalyptus* identificados como relacionados a genes de celulose sintase de parede secundária.

Grupos de ortologia			
AtCesA4	*AtCesA7*	*AtCesA8*	*Eucalyptus*
EUPE-XY-000-067-E09-MG.R	EUGR-XY-006-024-G03-RS.R	EUPE-XY-004-010-C09-RS.R	EUGL-XY-001-071-H08-AF.R
EUPE-XY-000-073-C06-MG.R	EUUR-XY-000-086-A03-BA.R	EUPE-XY-000-050-A03-MG.R	EUUR-XY-000-039-E08-BA.R
EUPE-XY-000-029-E07-MG.R	EUPE-XY-000-058-A04-MG.R	EUGL-XY-001-120-B11-AF.R	EUUR-XY-000-050-A09-BA.R
EUPE-XY-000-028-E09-MG.R	EUGR-XY-022-021-B10-BA.R	EUGL-XY-001-042-G06-AF.S	EUUR-XY-000-055-A02-BA.R
EUGL-XY-001-065-H05-AF.R	EUUR-XY-000-063-H08-BA.R	EUSP-FX-003-061-G01-BA.R	EUGL-XY-001-051-G02-RS.R
EUGR-XY-006-048-B09-RS.R		EUGL-XY-002-009-F11-CN.R	EUGL-XY-003-001-A11-CN.R
EUUR-XY-000-030-H09-BA.R		EUGR-XY-006-020-B03-RS.R	EUGR-XY-022-024-C06-BA.R
EUUR-XY-000-053-D07-BA.R			EUUR-XY-001-103-B12-BA.R
EUPE-XY-001-009-A03-RS.R			EUGL-XY-001-032-H12-RS.R
EUSP-FX-002-057-C10-CN.R			EUGL-XY-002-010-G03-CN.R
EUSP-FX-002-058-D10-CN.R			EUGL-XY-001-023-B03-RS.R
EUGR-PU-003-028-G10-MG.R			EUGR-XY-006-010-F11-RS.S
EUSP-FX-003-076-E06-BA.R			EUUR-XY-000-031-H07-BA.S
EUUR-XY-000-030-H10-BA.R			EUUR-XY-000-064-G07-BA.R
EUPE-XY-000-054-H02-MG.R			EUUR-XY-000-023-H02-BA.R
EUGR-XY-006-002-A01-RS.R			EUUR-XY-000-087-F02-BA.R
EUUR-XY-000-010-H03-BA.R			EUGL-XY-001-064-G03-CN.R
EUUR-XY-000-050-B07-BA.R			EUGL-XY-001-106-A01-AF.R
EUUR-XY-000-058-A04-BA.R			EUGL-XY-001-106-B01-AF.R
EUUR-XY-000-088-H08-BA.R			EUPE-XY-000-057-B11-MG.R
EUGR-SE-001-162-G09-GO.R			EUPE-XY-003-024-A01-CN.R
EUUR-XY-000-032-G05-BA.R			EUGL-XY-001-055-F10-AF.R
EUUR-XY-000-039-B08-BA.R			EUGL-XY-002-015-F05-CN.R
			EUUR-XY-000-034-F02-BA.R
			EUGL-XY-002-016-F05-CN.R

As seqüências pertencentes a cada grupo de ortologia (Tabela 3) foram então realinhadas. Isso foi feito para permitir uma análise comparativa mais refinada dentro de cada grupo, sem que houvesse, no alinhamento, a influência de outras seqüências não ortólogas. Após essa etapa de alinhamento, realizada com o ClustalX, foi observado que para os ESTs analisados houve a formação de somente um *cluster* para cada um dos grupos de ESTs considerados como ortólogos a *AtCesA4* e a *AtCesa8* e dois clusters de ESTs ortólogos a *AtCesA7*. Esses resultados sugerem que em *Eucalyptus* existe um único gene ortólogo a cada um dos genes *AtCesA4* e *AtCesA8*, e possivelmente o gene ortólogo a *AtCesA7* tenha passado por um evento de duplicação. Os alinhamentos obtidos para cada um dos quatro grupos de ortologia estão apresentados nos Anexos 1, 2, 3 e 4. Os resultados para cada grupo são apresentados separadamente a seguir.

5.2 Mineração e desenho de *primers* para genes ortólogos a *AtCesA4*

O alinhamento obtido para esse grupo de ortologia sugere a existência de apenas um potencial gene em *Eucalyptus* ortólogo ao gene *AtCesA4* de *A. thaliana*, relacionado à produção de celulose sintase de parede celular secundária (Anexo 1). Os *primers* foram desenhados considerando-se o polimorfismo representado no alinhamento e a posição do íntron (Figura 8), inferida pela comparação com a seqüência genômica do gene *AtCesA4*, através do programa EST2GENOME (Mott, 1997). O fragmento representado no alinhamento corresponde a uma região 3' do gene *AtCesA4* de *A. thaliana* e contém 951 pares de base, sendo 373 pares de bases pertencentes a um íntron flanqueado por duas regiões exônicas de 50 e 528 pares de base (Figura 9). Dois *primers* (CesA4.1 e CesA4.2) foram desenhados no intuito de se seqüenciar as regiões à montante e à jusante do potencial gene em *Eucalyptus*. Um outro par de *primers* (CesA4.3 e CesA4.4) foi desenhado a fim de se amplificar e/ou seqüenciar o maior éxon deste gene em *Eucalyptus* (Figura 9). A caracterização dos *primers* desenhados está apresentada na Tabela 4.

```
     :17731872-17737318   3909 TGTACTATTCCGGCCGTTTGTCTTCTCACCGGCAAGTTCATCATTCCTAC   3958
                               ||||| || || |||||||| |||||||| || || ||||| || || ||
EUUR-XY-000-088-H08-BA.R      1 TGTACAATACCTGCCGTTTGCCTTCTCACTGGGAAATTCATTATCCCCAC     50
```

Figura 8. continua...

```
      :17731872-17737318   3959 Ggtaag.....ttcagATCAACAATTTCGCAAGCATATGGTTCTTAGCTC   4366
                                |>>>>> 373 >>>>> |||  || || || |||||||||||||| || |
EUUR-XY-000-088-H08-BA.R     51 G..............CTCACTAACTTTGCGAGCATATGGTTCTTGGCCC     85

      :17731872-17737318   4367 TGTTCCTATCAATCATTGCAAC-GGCTATCTTAGAACTCCGGTGGAGTGG   4415
                                | |||||||| ||||| || || |||   ||| ||||| ||||||||||
EUUR-XY-000-088-H08-BA.R     86 TTTTCCTATCCATCATAGCCACTGGCGTGCTT-GAACTACGGTGGAGTGG    134

      :17731872-17737318   4416 TGTAAGCATCAATGACCTCTGGCGTAACGAGCAATTCTGGGTGATTGGCG   4465
                                ||| |||||| | |||   |||||||| || |||||||||| ||||| |
EUUR-XY-000-088-H08-BA.R    135 TGTCAGCATCGAGGACTGGTGGCGTAATGAACAATTCTGGGTCATTGGTG    184

      :17731872-17737318   4466 GTGTCTCAGCTCATCTCTTCGCTGTCTTTCAAGGTCTTCTTAAAGTTCTC   4515
                                | || || || || |||||||||| || ||||| || || || || ||
EUUR-XY-000-088-H08-BA.R    185 GCGTATCTGCACACCTCTTCGCTGTATTCCAAGGCCTCCTCAAGGTGCTT    234

      :17731872-17737318   4516 TTTGGTGTTGACACTAACTTCACCGTCACGTCAAAAGGAGCCAGTGACGA   4565
                                   || ||||| |||||||||| || ||  |||| | ||||   || |
EUUR-XY-000-088-H08-BA.R    235 GCCGGAGTTGATACTAACTTCACTGTTACAGCAAAGGCAGCC---GA-GG    280

      :17731872-17737318   4566 AGCCGATGAGTTTGGAGATCTTTACCTCTTTAAATGGACAACCCTTCTCA   4615
                                | | | ||||||||| || || ||||| || || ||||| ||||| ||||
EUUR-XY-000-088-H08-BA.R    281 A-CAG-TGAGTTTGGTGAACTCTACCTTTTCAAGTGGACTACCCTGCTCA    328

      :17731872-17737318   4616 TCCCTCCGACAACTCTCATTATTCTCAACATGGTCGGTGTTGTTGCCGGA   4665
                                | || || || ||||| || ||  | |||||||||||||| || |||||
EUUR-XY-000-088-H08-BA.R    329 TACCACCAACCACTCTAATAATCTTGAACATGGTCGGTGTCGTCGCCGGT    378

      :17731872-17737318   4666 GTTTCTGACGCTATTAACAATGGATATGG-TTCTTGGGGACCGCTTTTTG   4714
                                ||||| || || || |||||||||| || ||| | ||| |  ||  | |
EUUR-XY-000-088-H08-BA.R    379 GTTTCGGATGCCATAAACAATGGATACGGATTCGT-GGGCCNTCTGNTCG    427

      :17731872-17737318   4715 GGAAACTGTTTTTCGCGTTTTGGGTCATTGTTCATCTTTATCCTTTCCTC   4764
                                |||  ||  || |||  |||| ||| || | ||||| ||| ||||| |||
EUUR-XY-000-088-H08-BA.R    428 GGACGCT-CTTCTCGNCTTTT-GGTGATCG-TCATC-TTA-CCTTT-CTC    471

      :17731872-17737318   4765 AAAGGTTTGATGG   4777
                                   |||  ||||||
EUUR-XY-000-088-H08-BA.R    472 --AGGGCTGATGG    482
```

Figura 8. Posição relativa do íntron no alinhamento dos ESTs ortólogos ao gene *AtCesA4*, inferida pela utilização do programa EST2GENOME.

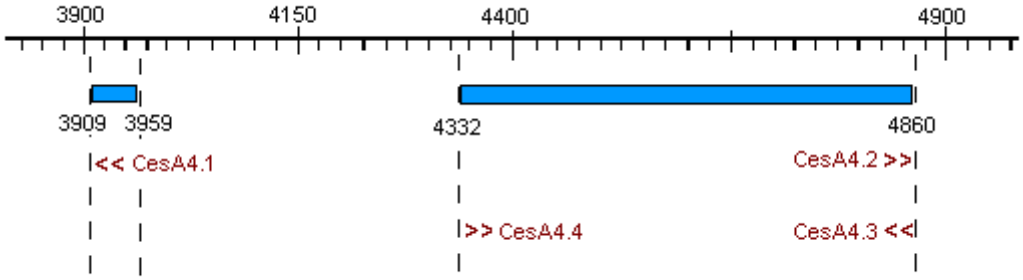


Figura 9. Representação esquemática dos resultados obtidos pelo programa EST2GENOME para a localização das regiões correspondentes a éxons e íntrons no alinhamento dos ESTs ortólogos ao gene *AtCesA4*. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA4*, as barras azuis representam o *contig* de ESTs de *Eucalyptus* ortólogas a esse gene. As setas indicam a localização e o sentido dos *primers* desenhados.

Tabela 4. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA4*.

nº do *primer*	nome	seqüência (5' → 3')	Temperatura de anelamento
1	CesA4.1	TTCCCAGTGAGAAGGCAAAC	56ºC
2	CesA4.2	TTGGTCCCTACTTCTCGCCT	56ºC
3	CesA4.3	AGGCGAGAAGTAGGGACCAA	56ºC
4	CesA4.4	TGAACTACGGTGGAGTGGTG	56ºC

5.3 Mineração e desenho de *primers* para genes ortólogos a *AtCesA7*

O alinhamento obtido neste caso (Anexo 2) sugeriu a possibilidade de que, em *Eucalyptus*, existem dois genes ortólogos ao gene *AtCesA7*, e portanto parálogos entre si. Essa hipótese é em parte corroborada pelos achados em *Populus tremuloides*, espécie para qual *PtrCesA2* é uma pequena família de genes composta por um ou dois membros que são homólogos à *AtCesA7* (Samuga & Joshi, 2002).

Para o primeiro gene potencialmente ortólogo a *AtCesA7* (aqui denominado de *CesA7.1*), três regiões em éxons, com 93, 353 e 278 pares de base, separadas por dois íntrons, foram inferidas com base na análise do alinhamento de aproximadamente 895 pares

de base (Anexo 2 e Figura 10). Neste caso, quatro *primers* foram desenhados (Figura 11). A caracterização destes *primers* está apresentada na Tabela 4.

```
       :c5741454-5736294   3427 TTAAAGAGGCAATCCATGTCATAAGCTGCGGTTATGAAGACAAGACTGAG   3476
                                |||  || || || ||||| ||  |||| || ||| | ||| ||||||||
EUGR-XY-022-001-A12-BA.R     29 TTATCGA-GC-ATTCATGTTATCCGCTGTGG-TAT-ANGAC-AGACTGAG     73

       :c5741454-5736294   3477 TGGGGAACTGAGgtaat....tgcagCTGGGTTGGATCTATGGCTCTATC   3602
                                ||||| |  |||>>>>> 90 >>>>> | || ||||| || || ||| ||
EUGR-XY-022-001-A12-BA.R     74 TGGGGCAAAGAG.............AT-GGATGGATATACGGTTCTGTC    108

       :c5741454-5736294   3603 ACAGAGGATATTTTGACGGGATTCAAGATGCATTGC-CGTGGATGGAGGT   3651
                                ||||| ||||| || || ||||| |||||||| ||| || || ||||  |
EUGR-XY-022-001-A12-BA.R    109 ACAGAAGATATCTTAACAGGATTTAAGATGCA-TGCTCGAGGTTGGATCT    157

       :c5741454-5736294   3652 CTATTTACTGCATGCCTAAGAGGCCTGCATTCAAAGGTTCAGCTCCTATT   3701
                                | |||||||| ||||||    | || |||||||| ||||| ||||| ||
EUGR-XY-022-001-A12-BA.R    158 CCATTTACTGTATGCCTCCACGCCCAGCATTCAAGGGTTCCGCTCCCATA    207

       :c5741454-5736294   3702 AATCTATCAGACAGGTTAAACCAGGTTTTGCGTTGGGCACTTGGATCGGT   3751
                                ||||| |||||  |  | ||||| ||| | || |||||| | || ||  |
EUGR-XY-022-001-A12-BA.R    208 AATCTTTCAGATCGTCTGAACCAAGTTCTTCGATGGGCATTGGGGTCCAT    257

       :c5741454-5736294   3752 TGAGATATTTTTCAGCCGGCACAGTCCTCTCTGGTATGGCTACAAAGGAG   3801
                                |||||| ||  | ||| ||||  ||||| | |||||||| ||| || | |
EUGR-XY-022-001-A12-BA.R    258 TGAGATTTTACTAAGCAGGCATTGTCCTATATGGTATGGTTAC-AATG-G    305

       :c5741454-5736294   3802 GCAAACTCAAGTGGCTTGAGCGTTTTGCTTATGCCAACACAACAATCTAC   3851
                                | | ||| |||||| | ||| | || || |||   || || |   | ||
EUGR-XY-022-001-A12-BA.R    306 G-AGACTGAAGTGGTTGGAGAGATTGGCATATATAAATACCATTGTGTAT    354

       :c5741454-5736294   3852 CCCTTCACATCTATACCACTTCTTGCCTACTGTATCCTTCCAGCCATCTG   3901
                                ||| |||| || || || ||  |||| || || || || || ||  | ||
EUGR-XY-022-001-A12-BA.R    355 CCCCTCACTTCAATCCCTCTGATTGCTTATTGCATTCTGCCTGCATTTTG    404

       :c5741454-5736294   3902 TCTCCTTACTGACAAATTCATCATGCCACCGgttag....ttcagATAAG   4017
                                 || |||||||  ||||| ||||| |||  |>>>>> 80 >>>>>|| ||
EUGR-XY-022-001-A12-BA.R    405 CCTTCTTACTGGAAAATTTATCATCCCAGAG.............ATTAG    440

       :c5741454-5736294   4018 CACATTTGCTAGTCTCTTCTTCA-TCTCACTGTTTATGTCGATCATTGTA   4066
                                ||  ||||||||| |  | |||| |||| || ||| | || ||  |||
EUGR-XY-022-001-A12-BA.R    441 CAACTTTGCTAGTAT-GTGTTCATTCTC-CTCTTTGTCTCCATTTTTGCT    488

  :c5741454-5736294   4067 ACGGGAATCTTGGAATTGAGATGGAGCGGAGTTAGCATTGAGGAATGGTG   4116
                                || |||||| | ||  | |||||||| ||||| || |||||||| |||||
EUGR-XY-022-001-A12-BA.R    489 ACCGGAATCCTTGAGCTCAGATGGAGTGGAGTGAGTATTGAGGATTGGTG    538
```

Figura 10. continua...

```
       :c5741454-5736294   4117 GAGAAACGAGCAATTCTGGGTCATTGGAGGAATCTCAGCTCATCTCTTTG   4166
                                ||| |  || ||  | |||||||| || ||    |||||||||| ||||
EUGR-XY-022-001-A12-BA.R    539 GAGGACTGAACAGGTTTGGGTCATCGGGGGCCCGTCAGCTCATCTTTTTG    588

       :c5741454-5736294   4167 CGGTTGTCCAAGGTCTCCTCAAA-ATCTTAGCAGGCATTGACACAAACTT   4215
                                  ||  | ||||| ||| | |||  |||| |  || || ||||| |  ||
EUGR-XY-022-001-A12-BA.R    589 ACGTGTTTCAAGGGCTCTTGAAAGTTCTTGGTTGGAATCGACACCACTTT    638

       :c5741454-5736294   4216 CACCGTC-ACATCAAAGGCAACAGATGATGAT--GACTTTGGAGAACTTT   4262
                                ||| ||| || |||||||   | || || |||  || || |  ||||  |
EUGR-XY-022-001-A12-BA.R    639 CACTGTCAACTTCAAAGGGCTCGGACGAAGATGGGAATTCGCGGAACCCT    688

       :c5741454-5736294   4263 ACGCATTCAAATGGAC   4278
                                | |   |||| || ||
EUGR-XY-022-001-A12-BA.R    689 A-G-TGTCAAGTGAAC    702
```

Figura 10. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs ortólogos ao gene *AtCesA7*, grupo denominado de *CesA7.1*, inferida pela utilização do programa EST2GENOME.

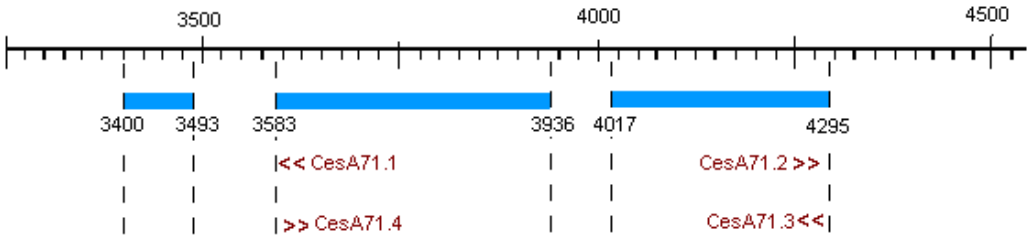


Figura 11. Representação esquemática do alinhamento realizado com o programa EST2GENOME para localização dos íntrons e éxons. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o primeiro grupo de ESTs de *Eucalyptus* ortólogos a esse gene (grupo *CesA7.1).* As setas indicam a localização e o sentido dos *primers* desenhados.

Tabela 5. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (*CesA7.1).*

nº do *primer*	nome	seqüência (5' → 3')	Temperatura de anelamento
1	CesA71.1	GCATGCATCTTAAACCCTGTT	56ºC
2	CesA71.2	CCAAGGTCTCCTCAAGGTTC	56ºC
3	CesA71.3	AACCTTGAGGAGACCTTGGA	56ºC
4	CesA71.4	AACAGGGTTTAAGATGCATGC	56ºC

A análise do alinhamento e do *contig* referente ao segundo grupo de ESTs, potencialmente ortólogo ao gene *AtCesA7* (aqui denominado de *CesA7.2*) detectou três regiões em éxons com 313, 356 e 350 pares de base, também separadas por dois íntrons (Figura 12). Para este gene, seis *primers* foram desenhados de modo a amplificar e seqüenciar regiões dentro do gene, bem como regiões que ultrapassavam os limites desses fragmentos, de maneira a seqüenciar a maior porção possível desse gene (Figura 13). A caracterização dos *primers* desenhados está apresentada na Tabela 6.

```
       :c5741454-5736294   3290 AGGAGCAGAAGGTGATAAGGAGCATTTGATGTCTGAAATGAACTTTGAGA   3339
                                ||||   || | ||| |||||||  |||||||| || ||||| |||||||
EUUR-XY-000-063-H08-BA.R     50 AGGAATGGATGATGACAAGGAGCTATTGATGTCCGAGATGAATTTTGAGA     99

       :c5741454-5736294   3340 AAACATTTGGGCAATCATCCATCTTTGTAACCTCAACTTTGATGGAAGAA   3389
                                | |  |||||||| ||  | || ||||| |||||||||||||||||| |
EUUR-XY-000-063-H08-BA.R    100 AGAAGTTTGGGCAGTCGGCAATTTTTGTTACCTCAACTTTGATGGAACAG    149

       :c5741454-5736294   3390 GGTGGTGTTCCTCCGTCATCAAGTCCTGCAGTGCTCCTTAAAGAGGCAAT   3439
                                ||||| || || || || |||||||| ||||  || || ||||| || ||
EUUR-XY-000-063-H08-BA.R    150 GGTGGCGTACCCCCTTCTTCAAGTCCAGCAGCACTGCTGAAAGAAGCCAT    199

       :c5741454-5736294   3440 CCATGTCATAAGCTGCGGTTATGAAGACAAGACTGAGTGGGGAACTGAGg   3488
                                |||||| || || || || |||||||| || ||||| ||||| || |||>
EUUR-XY-000-063-H08-BA.R    200 CCATGTTATCAGTTGTGGGTATGAAGATAAAACTGAATGGGGCACCGAG.    248

       :c5741454-5736294   3488 taat....tgcagCTGGGTTGGATCTATGGCTCTATCACAGAGGATATTT   3615
                                >>>> 90 >>>>> |||| ||||| ||||| ||||||||||||||||||
EUUR-XY-000-063-H08-BA.R    248 .............TTGGGCTGGATTTATGGTTCTATCACAGAGGATATCC    285

       :c5741454-5736294   3616 TGACGGGATTCAAGATGCATTGCCGTGGATGGAGGTCTATTTACTGCATG   3665
                                | || || || |||||||||||||||| |||||||| || || || |||
EUUR-XY-000-063-H08-BA.R    286 TAACAGGCTTTAAGATGCATTGCCGTGGTTGGAGGTCCATCTATTGTATG    335
```

Figura 12. continua...

```
       :c5741454-5736294   3666 CCTAAGAGGCCTGCATTCAAAGGTTCAGCTCCTATTAATCTATCAGACAG   3715
                                || ||||| || |||||||| || || ||||| |||||||| || || ||
EUUR-XY-000-063-H08-BA.R    336 CCAAAGAGACCAGCATTCAAGGGATCTGCTCCAATTAATCTTTCCGATAG    385

       :c5741454-5736294   3716 GTTAAACCAGGTTTTGCGTTGGGCACTTGGATCGGTTGAGATATTTTTCA   3765
                                | | || || ||  | |||||||||||||| || |||||||| ||||| |
EUUR-XY-000-063-H08-BA.R    386 GCTCAATCAAGTGCTTCGTTGGGCACTTGGTTCTGTTGAGATCTTTTTTA    435

       :c5741454-5736294   3766 GCCGGCACAGTCCTCTCTGGTATGGCTACAAAGGAGGCAAACTCAAGTGG   3815
                                |||  ||||||||  |||||||||| |||||||| || || || || |||
EUUR-XY-000-063-H08-BA.R    436 GCCACCACAGTCCAGTCTGGTATGGATACAAAGGGGGGAAGCTTAAATGG    485

       :c5741454-5736294   3816 CTTGAGCGTTTTGCTTATGCCAACACAACAATCTACCCCTTCACATCTAT   3865
                                || || || ||||| |||| ||||||||| ||||| || || || ||  |
EUUR-XY-000-063-H08-BA.R    486 CTCGAACGGTTTGCATATGTCAACACAACCATCTATCCTTTTACCTCCTT    535

       :c5741454-5736294   3866 ACCACTTCTTGCCTACTGTATCCTTCCAGCCATCTGTCTCCTTACTGACA   3915
                                |||||| |||||||||||||  ||||| ||||| ||  | || || ||||
EUUR-XY-000-063-H08-BA.R    536 ACCACTCCTTGCCTACTGTACTCTTCCGGCCATTTGCTTGCTCACAGACA    585

       :c5741454-5736294   3916 AATTCATCATGCCACCGgttag....ttcagATAAGCACATTTGCTAGTC   4031
                                | |||||||||||| | >>>>> 80 >>>>>|  ||||  ||||| ||||
EUUR-XY-000-063-H08-BA.R    586 AGTTCATCATGCCAGCA..............AATAGCA-TTTTGCCAGTC    620

       :c5741454-5736294   4032 TCTTCTTCAT-CTCACTGTTTATGTCGATCATTGTAACGGGAATCTTGGA   4080
                                | |||||| | ||| |  |||||||| ||| |||  ||  |  || | ||
EUUR-XY-000-063-H08-BA.R    621 T-TTCTTCTTGCTC-C--TTTATGTC-ATCTTTGCGAC-CGCCTC-T-GA    662

       :c5741454-5736294   4081 ATTGAGATGG-AG-CGGAGTTAGCATTGAGGAATGG   4114
                                | | || |||  | | ||  | | || | |||||||
EUUR-XY-000-063-H08-BA.R    663 ACT-AGGTGGCTGCCTGACCTGGGATGGGGGAATGG    697
```

Figura 12. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs ortólogos ao gene *AtCesA7*, grupo denominado de *CesA7.2*, inferida pela utilização do programa EST2GENOME.

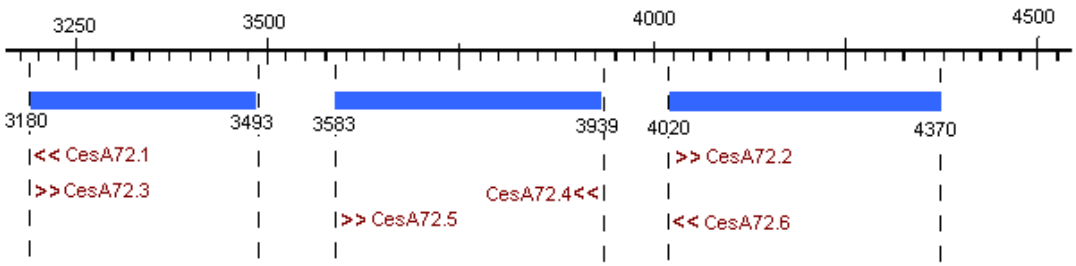


Figura 13. Representação esquemática do alinhamento realizado com o programa EST2GENOME para localização dos íntrons e éxons. A linha superior

demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o segundo grupo de ESTs de *Eucalyptus* ortólogos a esse gene (grupo *CesA7.2*). As setas indicam a localização e o sentido dos *primers* desenhados.

Tabela 6. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (*CesA7.2).*

nº do *primer*	nome	seqüência (5' → 3')	Temperatura de anelamento
1	CesA72.1	CCCCATTGGCACTATGCTTA	56ºC
2	CesA72.2	TTTGCGACCGGCATTCTT	56ºC
3	CesA72.3	TAAGCATAGTGCCAATGGGG	56ºC
4	CesA72.4	TGCTGGCATGATGAACTTGT	56ºC
5	CesA72.5	TAAGATGCATTGCCGTGGTA	56ºC
6	CesA72.6	AAGAATGCCGGTCGCAAA	56ºC

O alinhamento do grupo formado apenas por ESTs de *Eucalyptus* também foi construído (Anexo 3). A seqüência genômica de *AtCesA7* foi utilizada como referência neste caso. Foram observados dois grupos distintos de seqüências. Os *contigs* obtidos a partir do alinhamento de cada um destes grupos também foram comparados com os *contigs* dos demais grupos de ESTs (grupo *CesA7.1* e grupo *CesA72.*)e observou-se que as seqüências formadas apenas por ESTs apresentaram um elevado grau de similaridade com o segundo gene de *Eucalyptus* ortólogo ao gene *CesA7* de *Arabidopsis thaliana,* denominado de *CesA7.2,* se tratando na verdade de duas regiões distintas deste gene. A análise dos *contigs* obtidos para estes dois grupos denominados de EUC1 e EUC2, revelou que o primeiro se refere a uma região contendo 641 pares de base, que não inclui íntrons (Figuras 14 e 15). O segundo *contig* se refere a uma região que inclui 2 éxons, com 214 e 595 pares de base, e um íntron (Figura 16 e 17). As Tabelas 7 e 8 apresentam a caracterização dos *primers* desenhados nos fragmentos correspondentes a EUC1 e EUC2.

```
                1-3277   2814 ATCCCTCCAACCACTATCTTAATCATCAACCTTGTTGGAGTTGTTGCTGG   2863
                             |||||||||||| ||||| || |||||||||||| |||||||| ||||| ||
EUUR-XY-000-087-F02-BA.R      4 ATCCCTCCAACTACTATATTGATCATCAACCTAGTTGGAGTAGTTGCCGG     53

                1-3277   2864 AGTCTCAGATGCCATAAACAATGGGTACCAGTCATGGGGACCTCTATTCG   2913
                             | |||| ||||| || || ||||| |||||  |||||||||||| || |
EUUR-XY-000-087-F02-BA.R     54 AATCTCTGATGCTATTAATAATGGATACCAAGCATGGGGACCTCTGTTTG    103
```

Figura 14. continua...

```
                1-3277   2914 GGAAGCTCTTCTTTGCCTTCTGGGTGATTGTCCATCTCTACCCATTCCTC   2963
                             ||||  | ||||| ||||||||||||||| |||||||||| || |||||
EUUR-XY-000-087-F02-BA.R    104 GGAAATTGTTCTTCGCCTTCTGGGTGATTCTCCATCTCTATCCCTTCCTG    153

                1-3277   2964 AAAGGTCTAATGGGGAGGCAAAACAGGACACCGACTATTGTGGTTATATG   3013
                             ||||| || |||||||| || ||| | ||||| || || ||||| |||||
EUUR-XY-000-087-F02-BA.R    154 AAAGGGCTGATGGGGAGACAGAACCGTACACCAACAATAGTGGTCATATG    203

                1-3277   3014 GTCAGTGCTCCTGGCTTCCATCTTCTCCTTGCTTTGGGTCCGGATTGATC   3063
                             |||||| || |||||||||| |||||||||| |||||||| |||||||
EUUR-XY-000-087-F02-BA.R    204 GTCAGTACTACTGGCTTCCATATTCTCCTTGCTGTGGGTCCGAATTGATC    253

                1-3277   3064 CATTTGTGATGAAAACCAGGGGACCTGACACCAAGCAATGTGGACTCAAC   3113
                             | || ||| | || |||| ||| ||||| |||||| |||||||| ||||
EUUR-XY-000-087-F02-BA.R    254 CGTTCGTGCTAAAGACCAAGGGGCCTGATACCAAGAAATGTGGAATCAAT    303

                1-3277   3114 TGTTAAAAGTGTGTTTATTTTCTAGTTGA-TTTGTGCATCATATAGAAGA   3162
                             ||||  || | |    |||| ||  || | | |||||||| |    |||
EUUR-XY-000-087-F02-BA.R    304 TGTT--AATTCT-CAAATTTCCTGCTTCAGTGTGTGCATC-TCCTCAAG-    349

                1-3277   3163 TACATGTGCATCCTGCTTCTA-AGGAAATACGATATGCGATGTATAAACT   3211
                             |||| | | ||  | | || | | ||| |  | | | | | | |||  |
EUUR-XY-000-087-F02-BA.R    349 TACAAG-G-ATATTTCGTCCAGATGAATTTAG-TGTAC-A-G-ATATGCA    392

                1-3277   3212 GACTAAGATGGAGATGCTACAAGGAATAAAGTTAGAGTG-AAAATTTTGT   3260
                             | |  ||| | |  ||  | |  |||||| | ||||||| ||  ||||||
EUUR-XY-000-087-F02-BA.R    393 G-CACAGAAGTACTTG--AGATTGAATAAGGATAGAGTGCAATCTTTTGT    439

                1-3277   3261 GTA   3263
                             |||
EUUR-XY-000-087-F02-BA.R    440 GTA    442
```

Figura 14. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs denominado de EUC1, inferida pela utilização do programa EST2GENOME.

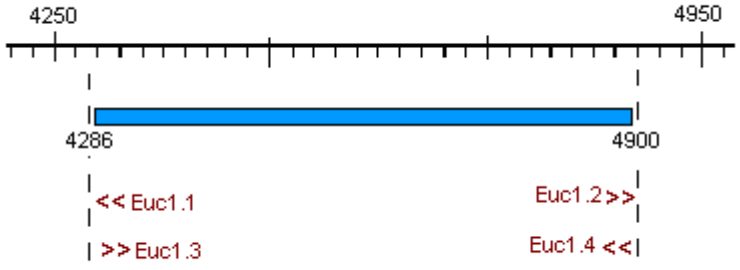


Figura 15. Representação esquemática do alinhamento realizado com o programa

EST2GENOME para localização dos íntrons e éxons. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o grupo de ESTs de *Eucalyptus* denominado de EUC1. As setas indicam a localização e o sentido dos *primers* desenhados.

Tabela 7. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (fragmento EUC1).

nº do *primer*	nome	seqüência (5' → 3')	Temperatura de anelamento
1	Euc1.1	CATCAGAGATTCCGGCAACT	56ºC
2	Euc1.2	GCAGCACAGAAGTGCTTGAG	56ºC
3	Euc1.3	AGTTGCCGGAATCTCTGATG	56ºC
4	Euc1.4	CTCAAGCACTTCTGTGCTGC	56ºC

```
       :c5741454-5736294   3742 TTGGATCGGTTGAGATATTTTTCAGCCGGCACAGTCCTCTCTGGTATGGC   3791
                                |||| || ||||| || ||||| ||||  ||||||||  ||||||||||
EUGL-XY-002-015-F05-CN.R       1 TTGGTTCTGTTGACATCTTTTTTAGCCACCACAGTCCAGTCTGGTATGGA     50

       :c5741454-5736294   3792 TACAAAGGAGGCAAACTCAAGTGGCTTGAGCGTTTTGCTTATGCCAACAC   3841
                                |||||||| || || || || ||||| || || ||||| |||| ||||||
EUGL-XY-002-015-F05-CN.R      51 TACAAAGGGGGGAAGCTTAAATGGCTCGAACGGTTTGCATATGTCAACAC    100

       :c5741454-5736294   3842 AACAATCTACCCCTTCACATCTATACCACTTCTTGCCTACTGTATCCTTC   3891
                                ||| ||||| || || || ||  ||||||| ||||||||||||  ||||
EUGL-XY-002-015-F05-CN.R     101 AACCATCTATCCTTTTACCTCCTTACCACTCCTTGCCTACTGTACTCTTC    150

       :c5741454-5736294   3892 CAGCCATCTGTCTCCTTACTGACAAATTCATCATGCCACCGgttag....   3932
                                | ||||| ||  | || || ||||| |||||||||||| | >>>>> 80
EUGL-XY-002-015-F05-CN.R     151 CGGCCATTTGCTTGCTGACAGACAAGTTCATCATGCCAGCA.........    191

       :c5741454-5736294   3932 ttcagATAAGCACATTTGCTAGTCTCTTCTTCATCTCACTGTTTATGTCG   4057
                                >>>>>|| ||||| ||||| ||||| ||||||||  | || ||||||||
EUGL-XY-002-015-F05-CN.R     191 .....ATTAGCACTTTTGCCAGTCTTTTCTTCATTGCTCTTTTTATGTCA    236

       :c5741454-5736294   4058 ATCATTG-TAACGGGAATCTTGGAATTGAGATGGAGCGGAGTTAGCATTG   4106
                                ||| |||  | ||| | |||| ||| |||| ||||| ||||| |||||||
EUGL-XY-002-015-F05-CN.R     237 ATCTTTGCGACCGGCATTCTT-GAACTGAGGTGGAGTGGAGTGAGCATTG    285

       :c5741454-5736294   4107 AGGAATGGTGGAGAAACGAGCAATTCTGGGTCATTGGAGGAATCTCAGCT   4156
                                |||||||||||||||| |||||||||||||||||| ||  | ||||||
EUGL-XY-002-015-F05-CN.R     286 AGGAATGGTGGAGAAATGAGCAATTCTGGGTCATTGGTGGTGTGTCAGCT    335
```

Figura 16. continua...

```
       :c5741454-5736294   4157 CATCTCTTTGCGGTTGTCCAAGGTCTCCTCAAAATCTTAGCAGGCATTGA   4206
                                ||  | ||||| | ||| |||||||||||| ||  | || || || ||  |
EUGL-XY-002-015-F05-CN.R    336 CACTTGTTTGCTGATGTACAAGGTCTCCTGAAGGTATTGGCTGGAATATA    385

       :c5741454-5736294   4207 CACAAACTTCACCGTCACATCAAAGGCAACAGATGATGATGACTTTGGAG   4256
                                ||| || ||||| || ||||| || ||  | ||||| || |||||||| |
EUGL-XY-002-015-F05-CN.R    386 CACTAATTTCACAGTGACATCCAAAGCGTCTGATGACGAAGACTTTGGTG    435

       :c5741454-5736294   4257 AACTTTACGCATTCAAATGGACAACACTGCTGATCCCTCCAACAACTGTC   4306
                                |||| ||||||||||||||||||||  || | |||||||||||| ||| |
EUGL-XY-002-015-F05-CN.R    436 AACTGTACGCATTCAAATGGACAACGTTGTTAATCCCTCCAACTACTATA    485

       :c5741454-5736294   4307 TTAATCATAAACATTGTTGGCGTTGTTGCAGGCATCTCAGATGCCATTAA   4356
                                || ||||| ||| | ||||| || ||||| || ||||| ||||| ||||
EUGL-XY-002-015-F05-CN.R    486 TTGATCATCAACCTAGTTGGAGTAGTTGCCGGAATCTCTGATGCTATTAC    535

       :c5741454-5736294   4357 CAATGGATATCAGTCTTGGGGACCTCTATTTGGTAAACTCTTCTTCTCCT   4406
                                 |||||||| ||  | ||| ||||||| ||||  ||| |  ||||| |||
EUGL-XY-002-015-F05-CN.R    536 TAATGGATACCAAGCATGGAGACCTCTGTTTGAGAAATTGATCTTCGCCT    585

       :c5741454-5736294   4407 TTTGGGTCATTGTTCATCTCTACCCATTCCTCAAAGGTCTGATGGGTAGA   4456
                                | || || ||| | |||||||| || || || ||| | |||||||| |||
EUGL-XY-002-015-F05-CN.R    586 TCTGAGTGATTCTCCATCTCTATCCCTTACTGAAAAGGCTGATGGGGAGA    635

       :c5741454-5736294   4457 CAGAACAGAACACCAACCATTGTGGTGATTTGGTCAGTGTTATTGGCATC   4506
                                || ||| | |||||||| || ||||| || ||||||||  || |||| ||
EUGL-XY-002-015-F05-CN.R    636 CACAACCGTACACCAACAATAGTGGTCATATGGTCAGTACTACTGGCTTC    685

       :c5741454-5736294   4507 TATCTTCTCTTTGCTTTGGGTAAGAATTGATCCTTTTGTGCTCAAGACCA   4556
                                 || | ||| |||||   |||   |||||||||  | ||||| ||||||
EUGL-XY-002-015-F05-CN.R    686 CATATACTCCTTGCTGCCGGTCCAAATTGATCCGCTCGTGCTTAAGACC-    735

       :c5741454-5736294   4557 AAGGACCTGACACTTCCAAGTGTGGCATCAACTGCTGAAGC-AAAAT   4602
                                |||| |||||  |   |||  |||| ||||| || | || |  ||||
EUGL-XY-002-015-F05-CN.R    735 AAGGGCCTGA-TCCCACAAAAGTGGAATCAATTG-TCAATCTTAAAT    779
```

Figura 16. Posição relativa dos íntrons e éxons no *contig* obtido do alinhamento do grupo de ESTs denominado de EUC2, inferida pela utilização do programa EST2GENOME.

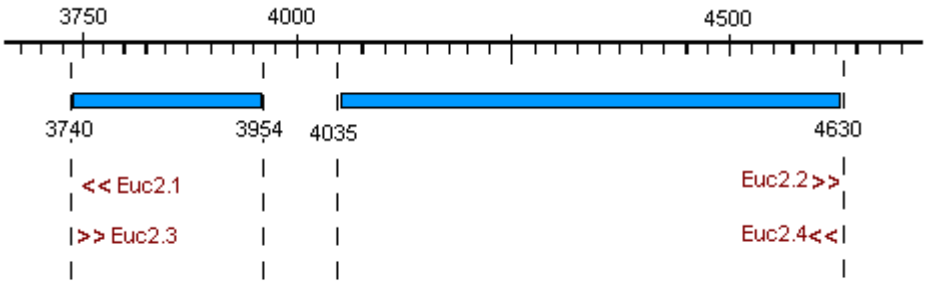


Figura 17. Representação esquemática do alinhamento realizado com o programa EST2GENOME para localização dos íntrons e éxons. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras azuis representam o *contig* obtido para o grupo de ESTs de *Eucalyptus* denominado de EUC2. As setas indicam a localização e o sentido dos *primers* desenhados.

Tabela 8. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA7* (fragmento EUC2).

nº do *primer*	nome	seqüência (5' → 3')	Temperatura de anelamento
1	Euc2.1	AGTAGGCAAGGAGTGGTAAGG	56ºC
2	Euc2.2	TTCCTGAAAGGGCTGATGG	56ºC
3	Euc2.3	CCTTACCACTCCTTGCCTACT	56ºC
4	Euc2.4	CCATCAGCCCTTTCAGGAA	56ºC

O alinhamento realizado com as seqüências homólogas a *AtCesA7* representantes do gene *CesA7.2* e os fragmentos EUC1 e EUC2 sugeriram que estes *contigs* se referem ao mesmo gene. A Figura 18 apresenta a disposição relativa dos *contigs* obtidos dos grupos *CesA7.2*, EUC1 e EUC2.

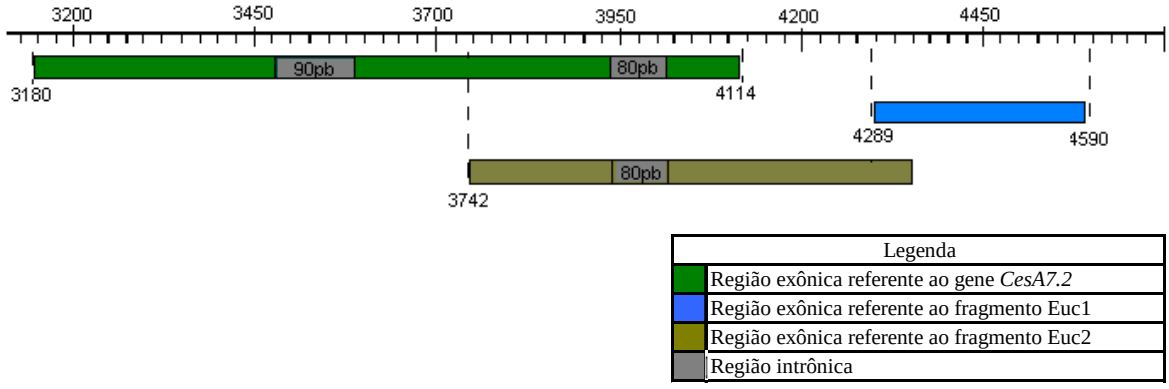


Figura 18. Representação esquemática dos três *contigs* relativos ao gene denominado *CesA7.2*. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA7*, as barras coloridas representam os *contigs* de ESTs de *Eucalyptus* ortólogos a esse gene.

Pela análise da Figura 18 é possível se inferir o motivo pelo qual algumas seqüências de ESTs, agora observadas como fazendo parte do gene *CesA7.2*, não formaram um *cluster* com nenhum dos genes de celulose sintase de *Arabidopsis thaliana* utilizados como referência nos primeiros alinhamentos (Tabela 1), mesmo tendo sido encontrados como homólogos aos genes *AtCesA* durante a mineração dos dados. Os tamanhos inferidos para os fragmentos denominados EUC1 e EUC2 são relativamente pequenos (aproximadamente 300pb e 540pb) quando comparados ao tamanho estimado para o fragmento denominado de *CesA7.2* (934pb). Além disso, a região de homologia entre os fragmentos do gene *CesA7.2* e EUC1 é nula e entre CesA7.2 e EUC2 é de somente 300 pares de base. Assim, é de se esperar que durante um alinhamento realizado com diversas seqüências de ESTs apresentando comportamentos similares, estas formem um grupo distinto.

5.4 Mineração e desenho de *primers* para genes homólogos a *AtCesA8*

O alinhamento realizado com o gene *AtCesA8* indicou que o *contig* obtido dos ESTs de *Eucalyptus*, neste caso, inclui três regiões em éxons, com 81, 514 e 161 pares de base cada, separadas por dois íntrons (Figura 19). Seis *primers* foram desenhados para este fragmento, sendo dois deles (CesA8.1 e CesA8.2) destinados ao seqüenciamento das

regiões à montante e à jusante do fragmento analisado. Outros quatro *primers* (CesA8.3, CesA8.4, CesA8.5 e CesA8.6) foram desenhados objetivando estudos de seqüenciamento e/ou amplificação de regiões dentro deste gene de celulose sintase em *Eucalyptus*, ortólogo ao gene *AtCesA8* (Figura 20). A caracterização dos *primers* obtidos está apresentada na Tabela 9.

```
     :c10316797-10312665   2243 GTTCAGTTCCCACAGAGATTTGATGGAATCGACAAGAGTGATCGATATGC   2292
                                || || ||||| || |||||||||||| || ||   || ||| || |||||
EUPE-XY-004-010-C09-RS.R      6 GTCCAATTCCCTCAAAGATTTGATGGCATTGATGCGAATGACCGGTATGC     55

     :c10316797-10312665   2293 TAACCGCAACATTGTTTTCTTCGATgtaag....ttcagGTTAATATGAG   2410
                                 ||| | ||||  |||||||| |||>>>>> 82 >>>>> | || ||||
EUPE-XY-004-010-C09-RS.R     56 CAACAGGAACACAGTTTTCTTTGAT..............ATCAACATGAA     91

     :c10316797-10312665   2411 AGGGCTTGATGGGATTCAAGGTCCAGTTTATGTTGGTACAGGTACTGTCT   2460
                                ||| || ||||| || || || ||||| ||||| || || ||   |||||
EUPE-XY-004-010-C09-RS.R     92 AGGTCTAGATGGAATCCAGGGCCCAGTATATGTGGGCACTGGATGTGTCT    141

     :c10316797-10312665   2461 TTAGAAGACAAGCACTTTACGGATACAGTCCACCTTCAAAACCGAGGATT   2510
                                |  | ||||| ||| | || || |  | |  |||  |    || ||| |
EUPE-XY-004-010-C09-RS.R    142 TCCGGAGACAGGCATTATATGGCT--A-TGAACC--C---CCCAAGGGTC    183

     :c10316797-10312665   2511 TTACCGCAATCTTCATCATCGTC-G-TGTTGCTGTCTAACCAAGAAGAAA   2558
                                 ||  ||  || | |  || ||| | || || |  ||  ||| |
EUPE-XY-004-010-C09-RS.R    184 CTA-AGC-GTCCT-AAGATGGTCAGCTG-TGATTGCTGCCCATGCTTTGG    229

     :c10316797-10312665   2559 CAACCTCAAGATCCTTCCGAGATTTACAAAGATGCAAAGCGAGAAGAACT   2608
                                |  || ||| |  || ||||  | | |  | | ||| || |   | |
EUPE-XY-004-010-C09-RS.R    230 CCGCCGCAAAAAACTCCCGAAGTATTC-TA-A-GCATAGTG-CCA-ATGG    274

     :c10316797-10312665   2609 TGATGCTGCAATCTTTAATCTTGGGGACCTAGACAACTACGATGAGTACG   2658
                                 |||||||| |      ||| |   |  | ||  || |  |||||  ||
EUPE-XY-004-010-C09-RS.R    275 GGATGCTGC-A-----GATC-T---G--CAAG-GAA-T-GGATGATGAC-    309

     :c10316797-10312665   2659 ACAGATCGATGCTGATTTCACAAACAAGCTTTGAGAAAACGTTTGGTCTC   2708
                                |  || | ||  |||| ||  | |  |  |||||||| | |||||| |
EUPE-XY-004-010-C09-RS.R    309 AAGGAGCTAT--TGATGTCCGAGATGAATTTTGAGAAGAAGTTTGGGCAG    356

     :c10316797-10312665   2709 TCTACGGTTTTCATCGAGTCT-ACTCTTATGGAGAATGGCGGTGTTCCCG   2757
                                ||  |  ||||  |  |   | ||| | |||||  | || || || |||
EUPE-XY-004-010-C09-RS.R    357 TCGGCAATTTTTGT-TACCNTCACTTTGATGGAACAGGGTGGCGTACCCC    405

     :c10316797-10312665   2758 ACTCTGTAAACCCGTCAACGCTCATCAAAGAAGCTATTCATGTCATTAGC   2807
                                  ||| | |  ||  || | ||  | |||||||| || ||||| || ||
EUPE-XY-004-010-C09-RS.R    406 CTTCTTTCAGTCCAGCAGCACTGCTGAAAGAAGCCATCCATGTTATCAGT    455
```

Figura 19. continua...

```
     :c10316797-10312665   2808 TGTGGATACGAAGAGAAAACTGAATGGGGAAAAGAAgtaag....cgcag   2843
                                |||||  ||  |||||  |||||||||||||  | |||>>>>> 73 >>>>>
EUPE-XY-004-010-C09-RS.R    456 TGTGGGTATGAAGA-TAAACTGAATGGGG-CACGAA.............    489

     :c10316797-10312665   2917 ATAGGATGGATTTACGGGTCGATCACCGAAGACATTTTGACGGGTTTCAA   2966
                                 | ||  ||||  |||  ||  ||  |||||  | ||   ||    ||  ||  ||  ||
EUPE-XY-004-010-C09-RS.R    490 TTGGGCTGGA-TTAAGGTTCTATCACAG-AG-GATACCCACAGGCTTTAA    536

     :c10316797-10312665   2967 GATGCATTGTCGTGGATGGAGGTCGATTTACTGTATGCCATTAAGACCAG   3016
                                 ||||||||  |  |||  ||||||||  |  |  |  ||  ||||||   ||||  |||
EUPE-XY-004-010-C09-RS.R    537 -ATGCATTG-CCTGGTTGGAGGTC-A-TAAATGGATGCCA-AAAGA-CAG    580

     :c10316797-10312665   3017 CATTTAAAGGATCTGCTCCAATCAATCTATCAGATAGGCTTCA-CCAGGT   3065
                                 |  ||  |  ||||||||  ||  ||   |  ||  ||   |  ||||   ||  |  |  ||
EUPE-XY-004-010-C09-RS.R    581 -A-TTCAGGGATCTGC-CC-AT-TA-CT-TC-CAAAGGC--CATCAATGT    620

     :c10316797-10312665   3066 TCTACGTTGGGCTCTTGGCTCGGTTGAGATCTT   3098
                                 ||   |  ||  |||  |||||  |   |||||  |  |||
EUPE-XY-004-010-C09-RS.R    621 TC--C-TT-GGC-CTTGG-T-TGTTGA-AACTT    645
```

Figura 19. Alinhamento e identificação de regiões intrônicas e exônicas para o gene *AtCesA8*, usando a seqüência genômica deste e um EST homólogo de *Eucalyptus*, através do programa EST2GENOME.

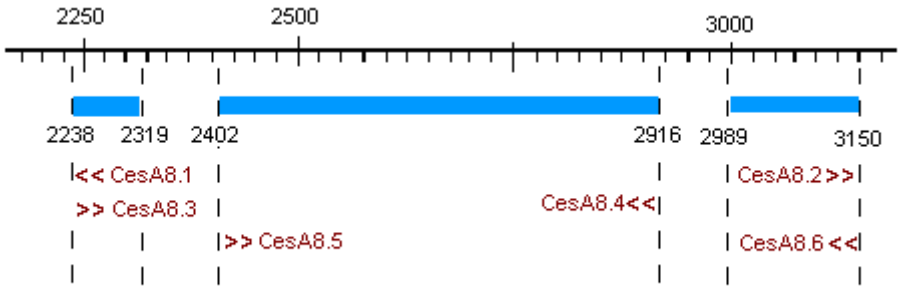


Figura 20. Representação esquemática dos resultados obtidos pelo programa EST2GENOME para a localização das regiões correspondentes a éxons e íntrons no alinhamento dos ESTs ortólogos ao gene *AtCesA8*. A linha superior demarcada representa a seqüência genômica de referência do gene *AtCesA8*, as barras azuis representam o *contig* de ESTs de *Eucalyptus* ortólogas a esse gene. As setas indicam a localização e o sentido dos *primers* desenhados.

Tabela 9. *Primers* desenhados para amplificação e/ou seqüenciamento do gene de celulose sintase de *Eucalyptus* ortólogo ao gene *AtCesA8*.

nº do *primer*	nome	seqüência (5' → 3')	Temperatura de anelamento
1	CesA8.1	ATGCAAGTCAATGCCGTCA	56ºC
2	CesA8.2	GCTCGAGGTTGGATCTCCA	56ºC
3	CesA8.3	ATGCAAGTCAATGCCGTCA	56ºC
4	CesA8.4	TCTTTGCCCCACTCAGTCTT	56ºC
5	CesA8.5	AAACTTGAGAGGTTTGGATGG	56ºC
6	CesA8.6	TGGAGATCCAACCTCGAGC	56ºC

Todos os *primers* destinados à amplificação e/ou seqüenciamento de regiões dentro dos genes de interesse foram desenhados de maneira que o tamanho esperado do fragmento a ser amplificado não ultrapassasse 1000 pares de bases. Além disso, os *primers* desenhados contêm um tamanho aproximado de 20 pares de base, a fim de se assegurar a especificidade dos mesmos.

5.5 Triagem da biblioteca de BACs para clones contendo fragmentos de *CesA7.2*

A triagem da biblioteca de BACs de *Eucalyptus grandis*, previamente construída no âmbito do projeto Genolyptus, foi realizada com os *primers* que flanqueavam regiões internas do gene *CesA7.2*. Inicialmente, três pares de *primers* foram utilizados para a PCR juntamente com o DNA do conjunto de *superpools*. Os três pares de *primers* (CesA72.3 e CesA72.4; EUC1.3 e EUC1.4; EUC2.3 e EUC2.4) foram escolhidos de modo que, dada a presença do gene *CesA7.2* em um dado BAC, um fragmento fosse amplificado. Assim, para cada par de *primers*, 22 reações foram feitas (21 *superpools* e um DNA controle). Em seguida à identificação dos *superpools* positivos, a *triagem* foi realizada dentro de cada um dos dez *pools* referentes a cada *pool* positivo. E finalmente, as reações foram feitas em cada placa da biblioteca que foi indicada como positiva. Todas as reações foram confirmadas em gel de agarose 1,5% corado com brometo de etídeo e o marcador utilizado foi o *Ladder 1Kb plus ou 100pb* (Invitrogen).

5.6 Triagem da biblioteca de BACs com os *primers* CesA72.3 e CesA72.4

Na Figura 21, que mostra a confirmação das amplificações do DNA dos superpools para os *primers* CesA72.3 e CesA72.4, é possível ver que tal amplificação ocorreu para três dos 21 *superpools* (*superpools* de número 01, 17 e 21). A última banda do gel (da esquerda para direita) se refere à amplificação do DNA do controle positivo. O tamanho médio dos insertos amplificados com estes *primers* foi superior a 800pb. O resultado da amplificação deste fragmento em indivíduos de diferentes espécies de *Eucalyptus* é apresentado na Figura 22. Observa-se que o fragmento amplificado por esse par de *primers* possui um tamanho aproximado de 1000 pares de base.

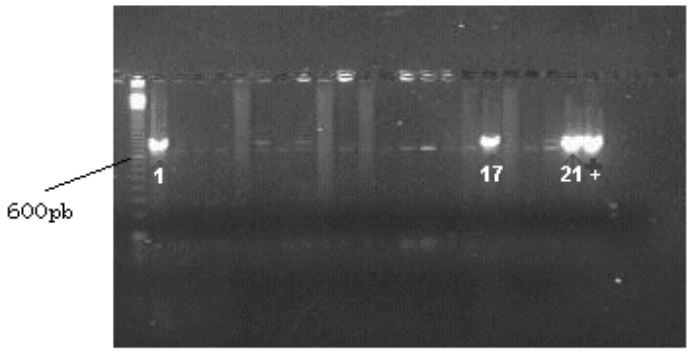


Figura 21. Confirmação da amplificação de fragmento do gene *CesA7.2* em *superpools*, utilizando os primers CesA72.3 e CesA72.4 (gel de agarose 1,5%) (1 – *superpool* 1; 17 – *superpool* 17; 21 – *superpool* 21 e + - controle positivo).

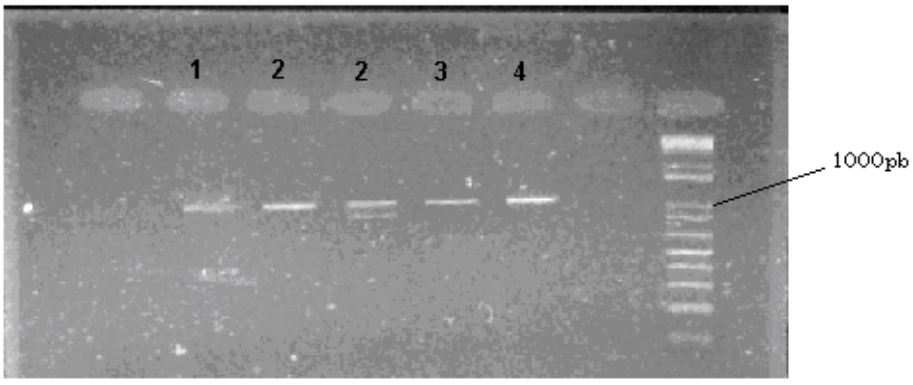


Figura 22. Confirmação da amplificação de fragmento do gene *CesA7.2* em indivíduos de diferentes espécies de *Eucalyptus* (1- *E. camaldulensis;* 2- *E. dunni;* 3- híbrido *E. urophylla e E. globulus;* 4- *E. grandis*), utilizando os primers CesA72.3 e CesA72.4 (gel de agarose 1,5%).

Para cada um dos três *superpools* positivos onze reações de amplificação foram feitas, sendo dez correspondentes aos *pools* e uma referente ao controle positivo. Para o *superpool* 01, dois pools foram confirmados como positivos (*pools* de número 002 e 003, correspondentes às placas EUGR-BC-002-002 e EUGR-BC-002-003). Para o *superpool* 17, apenas um *pool* foi positivo (164), assim como para o *superpool* 21 (206). Assim, quatro placas da biblioteca de BACs contêm o fragmento amplificado pelos *primers* CesA72.3 e CesA72.4. As reações de PCR foram então realizadas em cada uma dessas placas para a completa identificação dos clones positivos. A placa 02 teve o clone 11D como positivo. Na placa 03 foram identificados quatro clones positivos (4B, 4C, 5C e 6E), sugerindo a ocorrência de contaminação na placa estoque utilizada. Na placa 164 somente um clone foi positivo para esse par de *primers*: 11E. Para a placa 206 cinco clones foram identificados como positivos: 1B, 1C, 2A, 2B e 2C, novamente sugerindo a ocorrência de contaminação. Um resumo da identificação dos clones positivos para o fragmento flanqueado pelos *primers* CesA72.3 e CesA72.4 é apresentado na Tabela 10.

Tabela 10. Relação dos clones positivos para o fragmento do gene *CesA7.2*, identificados pela amplificação com os *primers* CesA72.3 e CesA72.4.

Superpool	*Pool* (EU-GR-002-XXX)	Localização na placa	Clone
01	002	11D	EUGR-BC-002-002-11D
	003	4B	EUGR-BC-002-003-04B
		4C	EUGR-BC-002-003-04C
		5C	EUGR-BC-002-003-05C
		6E	EUGR-BC-002-003-06E
17	164	11E	EUGR-BC-002-164-11E
21	206	1B	EUGR-BC-002-206-01B
		1C	EUGR-BC-002-206-01C
		2A	EUGR-BC-002-206-02A
		2B	EUGR-BC-002-206-02B
		2C	EUGR-BC-002-206-02C

5.7 Triagem da biblioteca de BACs com os *primers* EUC1.3 e EUC1.4

A triagem realizada com o par de *primers* EUC1.3 e EUC1.4 identificou como positivos os mesmos três *superpools* identificados pelos primers CesA72.3 e CesA72.4 (*superpools* de número 01, 17 e 21).

As reações de PCR realizadas a partir dos *pools* também identificaram os mesmos *pools* positivos que haviam sido identificados para o par de *primers* CesA72.3 e CesA72.4. Estes resultados confirmam que ambos os fragmentos estão localizados em regiões muito próximas no genoma de *Eucalyptus* e provavelmente correspondem ao mesmo gene, corroborando a estrutura esquematizada na Figura 18. Com a abertura dos *pools* para a completa identificação dos clones positivos, observou-se, no entanto, que alguns clones foram coincidentes e outros não. Embora, como comentado anteriormente, a possibilidade de ocorrência de contaminantes deve ser considerada, estes resultados sugerem que essas duas regiões possam não ter sido clonadas juntas em alguns clones durante a construção da biblioteca. Um exemplo disso é a amplificação realizada na placa 206 (Figura 23). O clone 3B foi positivo para o par de *primers* EUC1.3 e EUC1.4 e negativo para o par de *primers* CesA72.3 e CesA72.4. Enquanto que o inverso ocorreu para o clone EUGR-BC-002-206-2C. É importante considerar, entretanto, que dada a estrutura esquematizada na Figura 18, resultados desta natureza podem ocorrer. O resumo dos clones declarados positivos pela utilização dos *primers* EUC1.3 e EUC1.4 é apresentado na Tabela 11.

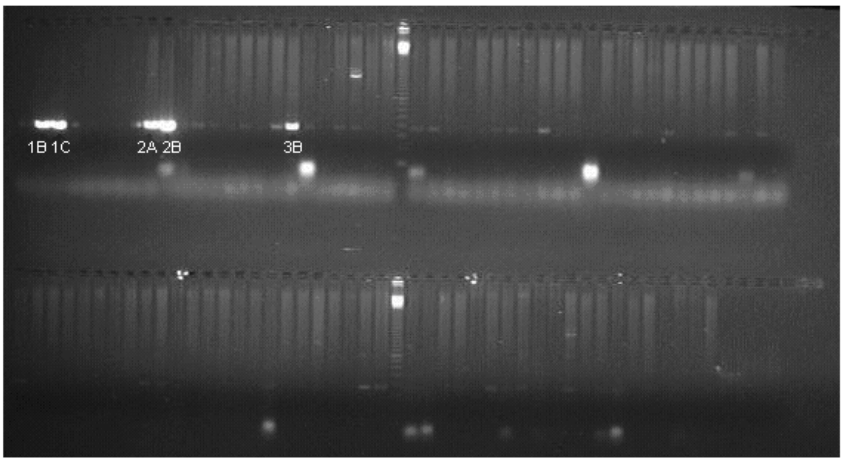


Figura 23. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC1.3 e EUC1.4 em clones da placa EUGR-BC-002-206 (gel de agarose 1,5%). Figura evidenciando posição relativa dos clones na placa.

Tabela 11. Relação dos clones positivos para o fragmento do gene *CesA7.2*, identificados pela amplificação com os *primers* EUC1.3 e EUC1.4.

Superpool	*Pool* (EU-GR-002-XXX)	Localização na placa	Clone
01	002	1H	EUGR-BC-002-002-01H
		2H	EUGR-BC-002-002-02H
	003	6E	EUGR-BC-002-003-06E
17	164	11E	EUGR-BC-002-164-11E
21	206	1B	EUGR-BC-002-206-01B
		1C	EUGR-BC-002-206-01C
		2A	EUGR-BC-002-206-02A
		2B	EUGR-BC-002-206-02B
		3B	EUGR-BC-002-206-03B

O resultado das amplificações realizadas em clones *Eucalyptus grandis* da placa 206 (Figura 24) permite sugerir que o tamanho estimado para este amplicon no genoma de *Eucalyptus* é de aproximadamente 600 pares de base. Esta estimativa está próxima ao tamanho esperado do fragmento com base nos alinhamentos realizados entre os ESTs de *Eucalyptus* e a seqüência genômica de *AtCesA7* (Anexo 3), que é de 614 pares de base.

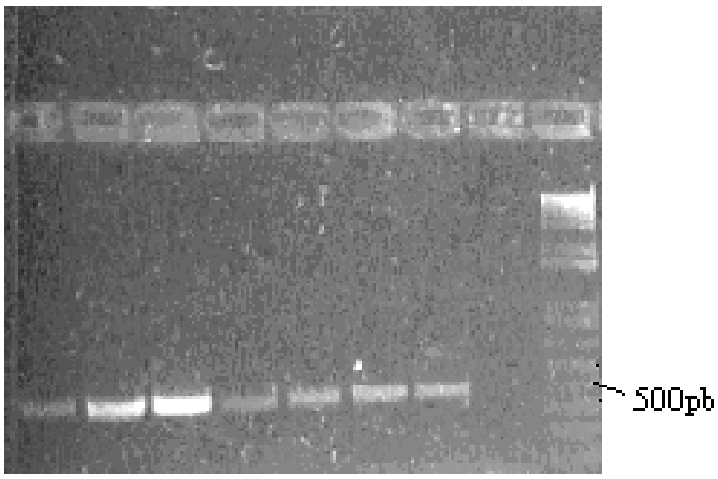


Figura 24. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC1.3 e EUC1.4 em clones da placa 206 (gel de agarose 1,5%). (As bandas no gel se referem, da direita para a esquerda, aos clones da placa 206 1B, 1C, 1C, 2A, 2A, 2B e 3B, respectivamente)

5.8 Triagem da biblioteca de BACs com os *primers* EUC2.3 e EUC2.4

Dos três pares de *primers* selecionados para a etapa de triagem da biblioteca de BACs, este foi o que detectou um maior número de clones positivos (15). Quatro *superpools* foram identificados com positivos (01, 10, 17 e 21), sendo três deles coincidentes com aqueles detectados pelos outros pares de *primers*, como mostrado na Figura 25. Os *pools* identificados como positivos para os *superpools* 01, 17 e 21 foram os mesmos identificados para os outros pares de *primers*. Para o *superpool* 10, o *pool* positivo foi o 098. Um total de 15 clones foi identificado como representantes do fragmento amplificado pelo par de *primers* EUC2.3 e EUC2.4 (Tabela 12). A placa 206 foi aquela em que um maior número de clones foram identificados como positivos (10), novamente sugerindo a ocorrência de contaminação (Figura 26). Estes dados corroboram novamente a hipótese formulada na Figura 18. Na figura 27 é possível observar a amplificação desse fragmento de aproximadamente 600pb em diferentes espécies de *Eucalyptus.*

A contaminação, que provavelmente devem ter ocorrido durante o processo de armazenamento das placas de BAC em ultra freezer a -80ºC, foi sugerida com base na redundância dessa biblioteca (4 vezes). A probabilidade de ocorrência de mais de 4 clones positivos na biblioteca dacai à medida que o número destes aumenta (Figura 28). Com base na Figura 28 é possivel perceber que a ocorrência de 15 clones positivos na biblioteca de BACs é um fato bastante improvável, porém as possíveis causas de uma provável contaminação devem ser melhor avaliadas em estudos posteriores.

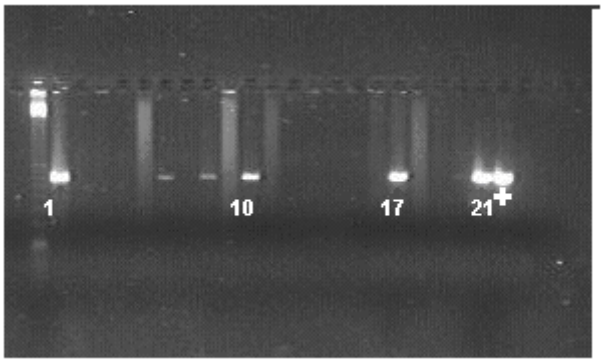


Figura 25. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC2.3 e EUC2.4 nos *superpools* (gel de agarose 1,5%). (1 – *superpool* 1; 10 – *superpool* 10; 17 – *superpool* 17; 21 – *superpool* 21 e + - controle positivo).

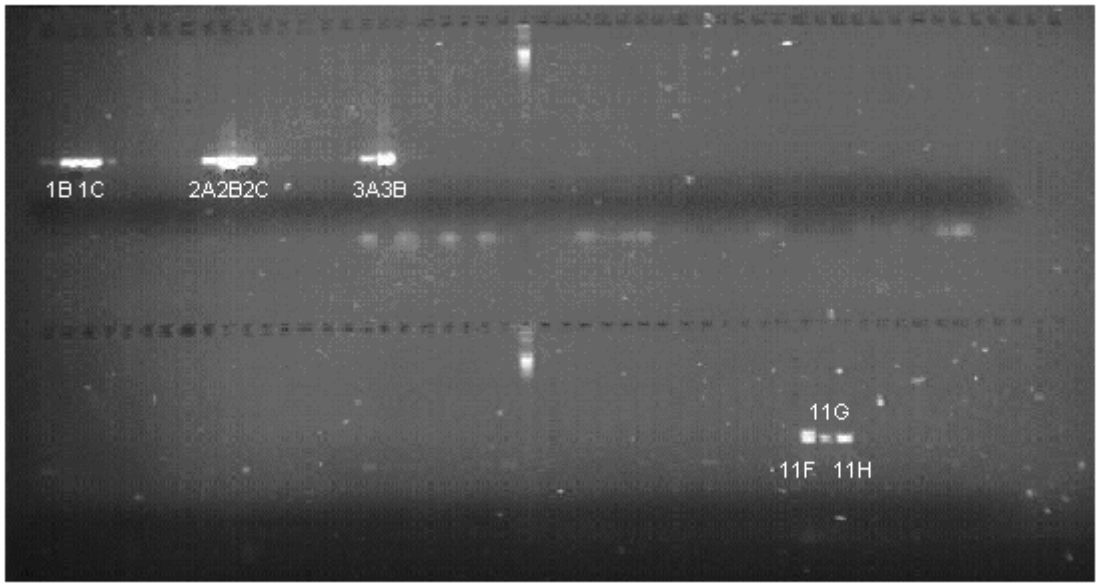


Figura 26. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC2.3 e EUC2.4 em clones da placa 206 (gel de agarose 1,5%). Figura evidenciando a posição relativa dos clones na placa.

Tabela 12. Identificação dos clones positivos pela amplificação com os *primers* EUC2.3 e EUC2.4.

Superpool	*Pool* (EU-GR-002-XXX)	Localização na placa	Clone
01	002	1H	EUGR-BC-002-002-01H
		2H	EUGR-BC-002-002-02H
	003	6E	EUGR-BC-002-003-06E
10	098	7B	EUGR-BC-002-098-07B
17	164	11E	EUGR-BC-002-164-11E
21	206	1B	EUGR-BC-002-206-01B
		1C	EUGR-BC-002-206-01C
		2A	EUGR-BC-002-206-02A
		2B	EUGR-BC-002-206-02B
		2C	EUGR-BC-002-206-02C
		3A	EUGR-BC-002-206-03A
		3B	EUGR-BC-002-206-03B
		11F	EUGR-BC-002-206-11F
		11G	EUGR-BC-002-206-11G
		11H	EUGR-BC-002-206-11H

O tamanho do fragmento amplificado pela utilização destes *primers* em diferentes espécies de *Eucalyptus* foi de aproximadamente 650 pares de base, como pode ser observado na Figura 27.

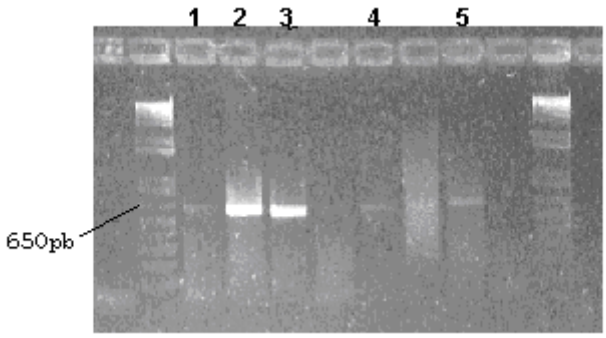


Figura 27. Confirmação da amplificação de fragmentos pela utilização dos *primers* EUC2.3 e EUC2.4 em indivíduos de diferentes espécies de *Eucalyptus* (1 e 2- *E. grandis;* 3- *E. urophylla*; 4- *E. dunni*; 5- *E. globulus*) (gel de agarose 1,5%).

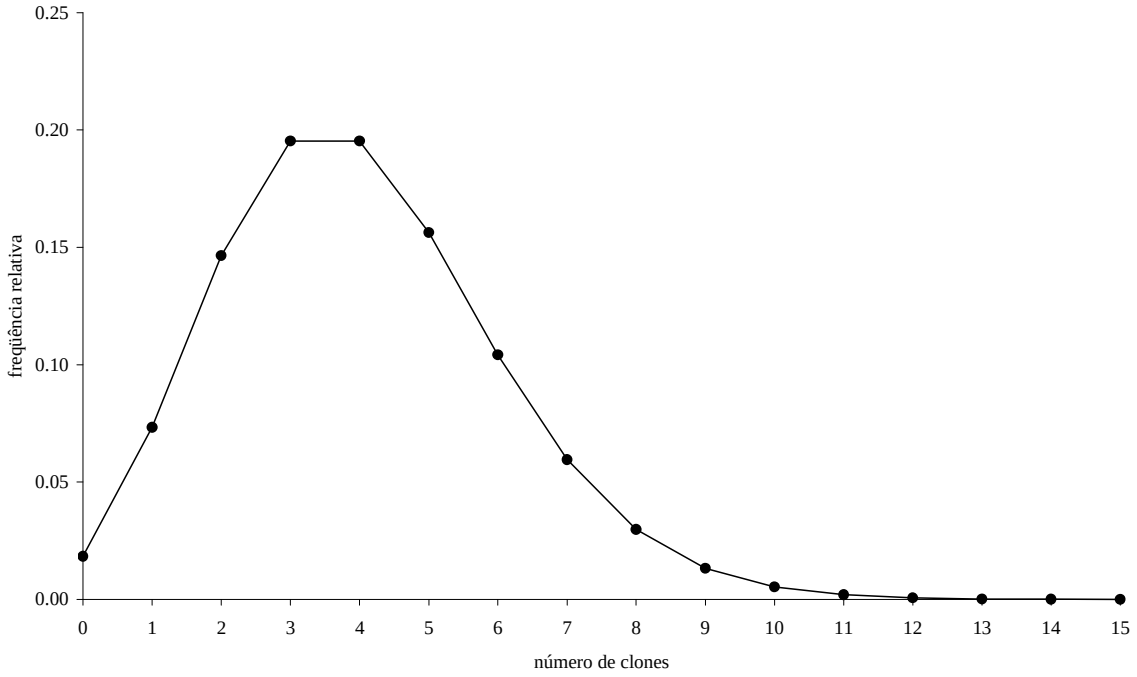


Figura 28. Freqüência da ocorrência de clones positivos na biblioteca de BACs construída para *Eucalyptus grandis.*

5.9 Confirmação da amplificação de fragmentos diretamente nos clones com os *primers* CesA72.5 e CesA72.6

A triagem na biblioteca dos BACs não foi realizada para o par de *primers* 72.5 e 72.6. No caso desses presume-se que o fragmento por eles amplificado esteja localizado nos mesmos clones que tiveram fragmentos amplificados para o par 72.3 e 72.4. Assim, o

teste para tais amplificações foi feito diretamente nos clones: EUGR-BC-002-002-11D, EUGR-BC-002-003-04B, EUGR-BC-002-003-04C, EUGR-BC-002-003-05C, EUGR-BC-002-003-06E, EUGR-BC-002-164-11E, EUGR-BC-002-206-01B, EUGR-BC-002-206-01C, EUGR-BC-002-206-02A, EUGR-BC-002-206-02B, EUGR-BC-002-206-02C (Figura 29). A fim de se aumentar a robustez das análises a amplificação neste caso foi realizada com temperatura de anelamento de 50ºC, inferior à temperatura de anelamento estimada inicialmente para os *primers* (56ºC). A diminuição da temperatura de anelamento dos *primers* diminui a especificidade dos mesmos durante a amplificação dos fragmentos, gerando diversas bandas que podem ser visualizadas no gel abaixo. Porém, apesar da inespecificidade, essa estratégia garante a amplificação dos fragmentos. Testes posteriores realizados para verificar a temperatura ideal de anelamento desses *primers* confirmaram a temperatura ideal em torno de 56ºC.

Assim, os clones cuja amplificação foi confirmada para os *primers* CesA72.5 e CesA72.6 foram: EUGR-BC-002-003-4C, EUGR-BC-002-003-5C, EUGR-BC-002-003-6E, EUGR-BC-002-164-11E e EUGR-BC-002-206-2B. Na Figura 29 é possível se observar que o tamanho dos fragmentos produzidos (de modo estringente) está em torno de 450pb, de acordo com o tamanho do fragmento esperado à partir das observações dos alinhamentos realizados com a seqüência de referência de *AtCesA7* de *A. thaliana* (437 pares de base).

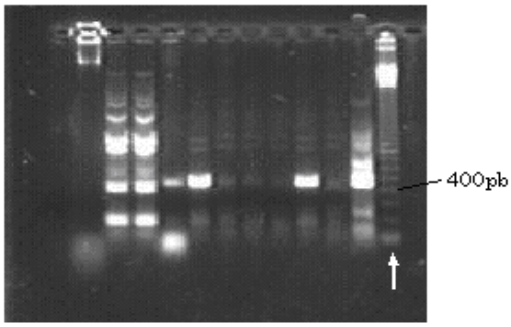


Figura 29. Fragmentos obtidos por amplificação de clones positivos, utilizando os *primers* CesA72.5 e CesA72.6. A penúltima canaleta (indicada pela seta) se refere ao controle positivo, seguida do *ladder* de 100pb (Invitrogen Technologies).

5.10 Isolamento do BACs por lise alcalina e amplificação dos fragmentos

A fim de se obter a seqüência genômica do gene *CesA7.2* de *Eucalyptus grandis*, alguns dos clones identificados como positivos foram selecionados para isolamento por lise alcalina. Seis clones BAC amplificados com pelo menos um dos quatro pares de *primers* (CesA72.3 e CesA72.4, CesA72.5 e CesA72.6, EUC1.3 e EUC1.4, EUC2.3 e EUC2.4) foram selecionados para essa fase do trabalho: EUGR-BC-002-002-01H, EUGR-BC-002-002-02H, EUGR-BC-002-002-11D, EUGR-BC-002-003-04C, EUGR-BC-002-003-06E e EUGR-BC-002-206-02B. Após o isolamento e lavagem com etanol 70%, o DNA dos clones, ressuspendido em 30μL de água ultra-pura foi quantificado em gel de agarose 1,5% e posteriormente submetido às reações de PCR com 45 ciclos a fim de que todo o *primer* disponível fosse utilizado durantes as reações. Os produtos de PCR foram então quantificados em gel de agarose 1,5%. Um exemplo dessas quantificações está mostrado na Figura 30. O objetivo geral dessa quantificação foi observar se os fragmentos foram amplificados antes de se realizar a etapa seguinte de seqüenciamento.

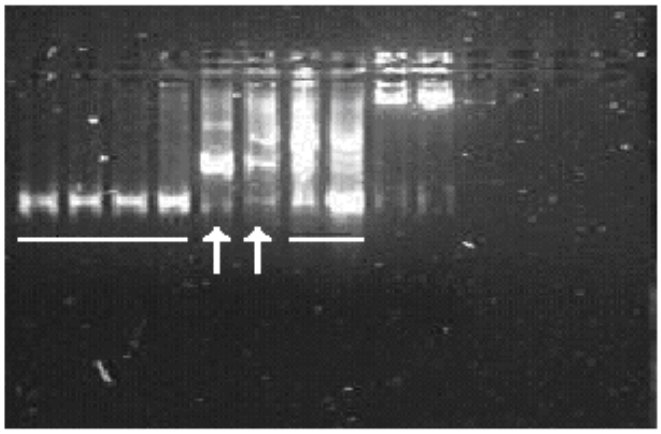


Figura 30. Visualização dos produtos de PCR utilizando o DNA dos clones selecionados e dois pares de *primers*: EUC2.3 e EUC2.4 (linha branca), e CesA72.3 e CesA72.4 (seta branca). As duas últimas canaletas correspondem a marcadores de massa molecular (50ng e 100ng de DNA do fago λ, respectivamente).

5.11 Seqüenciamento de parte do gene *CesA7.2* de *Eucalyptus grandis*

Uma vez seqüenciados, os fragmentos do gene *CesA7.2* amplificados com os *primers* descritos acima foram alinhados e montados com o uso do programa Staden Package. A montagem destes fragmentos gerou dois *contigs* de tamanhos distintos. Os *contigs* foram submetidos a uma análise utilizando o programa *blast* contra o banco de dados do GenBank (NR, acesso em setembro de 2006) para confirmação de sua identidade com genes de celulose sintase de outras espécies. Com base nos resultados do *blast* foi possível se observar a existência de uma seqüência de *Eucalyptus grandis* quase idêntica aos dois *contigs* obtidos neste trabalho. Esta seqüência, um cDNA completo do gene de celulose sintase denominado CesA3, é proveniente de um trabalho realizado por Alexander A. Myburg e Martin Ranik submetida ao GenBank durante o decorrer deste trabalho (número de acesso: DQ014507) (Ranik & Myburg, 2006). A identidade entre os dois *contigs* e o cDNA de *Eucalyptus* depositado no GenBank foi maior do que a identidade encontrada entre estes *contigs* e as seqüências disponíveis de *AtCesA7* (Contig1 – 89% de identidade com *AtCesA7* e 100% de identidade com *CesA3* de *E. grandis;* Contig 2 - 80% de identidade com *AtCesA7* e 98% de identidade com *CesA3* de *E. grandis*. A Figura 31 apresenta as seqüências obtidas para os dois *contigs*, utilizando a seqüência depositada no GenBank de *Eucalyptus grandis CesA3* como referência.

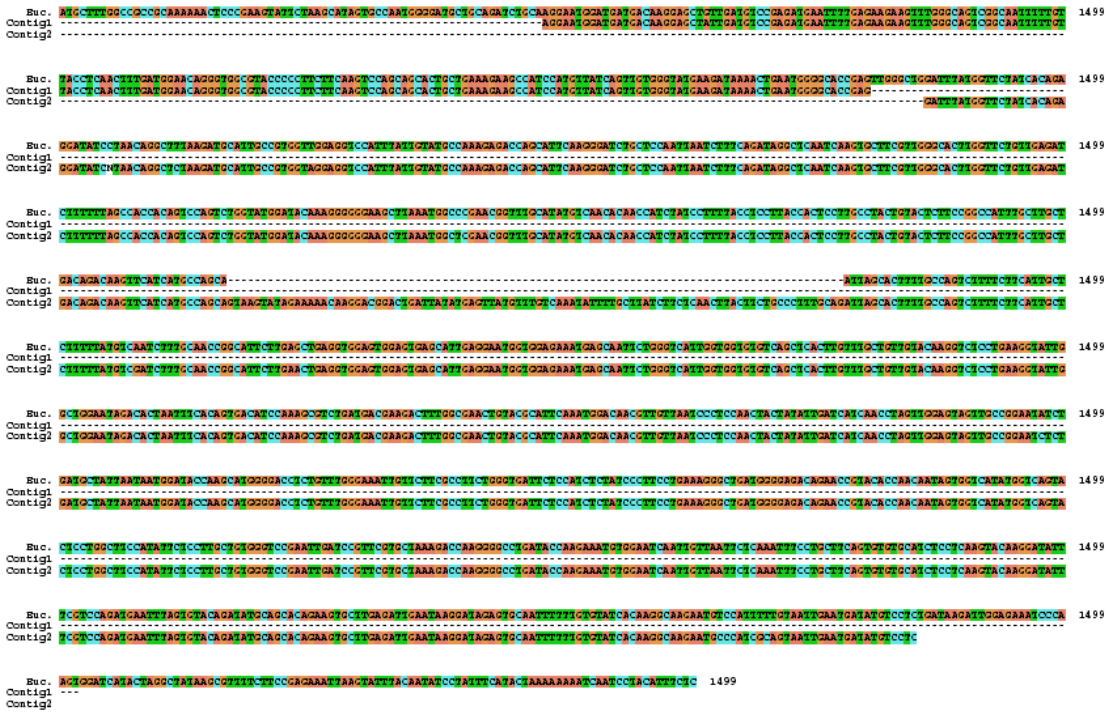


Figura 31. *Contigs* obtidos pelo seqüenciamento parcial do gene *CesA7.2* de *Eucalyptus*

grandis. Parte da seqüência depositada no *GenBank* como de *CesA3* de *Eucalyptus grandis* é mostrada como referência.

É possível observar na Figura 31 que o primeiro *contig*, formado por 199 pares de bases, está a 8 pares de bases de distância do segundo *contig*, formado por 1197 pares de bases. Este segundo *contig* inclui um íntron, cuja seqüência é inédita na literatura. É importante destacar que a seqüência do gene denominado de *CesA3* de *Eucalyptus grandis* depositada no *GenBank* é uma seqüência de cDNA, enquanto que os *contigs* obtidos neste trabalho são formados por seqüência genômicas de *Eucalyptus grandis.* A existência do íntron no *contig* 2 foi confirmada pelo EST2GENOME, que indicou a presença de um íntron de 92 pares de bases.

A estrutura do gene *AtCesA7* foi previamente elucidada na literatura, (Figura 32) os genes dessa família seguem um padrão, todos possuem um domínio do tipo dedo de zinco na região amino-terminal da proteína, regiões variáveis distribuídas pela proteína (ricas em aminoácidos ácidos) e oito domínios transmembrana, sendo seis deles localizados na região carboxi-terminal da proteína. As regiões variáveis foram denominadas por Ranik e Myburg (2006) de regiões específicas de classe, pois a variabilidade que ocorre é entre genes parálogos. Entre os domínios transmembrana existe uma região denominada de domínio globular e parece estar relacionada com a formação de um *loop* no citoplasma. Dentro desse domínio existem motivos conservados que são indicadores de glicosiltransferases (domínios A e B). O domínio A possui um conjunto de aminoácidos aspartato que é responsável pela ligação ao substrato UDP-glucose. Tais motivos são encontrados tanto em proteínas processivas (que catalizam a adição de muitos resíduos de açúcar à cadeia crescente) como em não-processivas (catalizam a adição de apenas um resíduo de açúcar). O domínio B consiste em um resíduo de aspartato (D) e um motivo conservado (QxxRW). Esses motivos fazem parte do sítio catalítico da proteína e estão presentes apenas em proteínas processivas (Richmond, 2000).

À partir da região aqui seqüenciada do gene *CesA7.2* de *Eucalyptus* é possível predizer uma proteína que possui homologia deste com a região carboxi-terminal da proteína CESA de *Eucalyptus* descrita por Ranik & Myburg (2006). Na Figura 33 pode-se observar que o fragmento seqüenciado abrange uma pequena porção da região específica de classe (traço vermelho) e todo o resto da protína pertence à região carboxi-terminal conservada. Dentro desta última é possível observar os motivos conservados que fazem parte do domínio B (D, QxxRW), seis domínios transmembrana da região carboxi-

terminal (traços azuis). As setas pretas indicam a existência de duas subtituições não-sinônimas. O *gap* de 8 pares de base formado entre os *contigs* 1 e 2 (Figura 31) está representado pelas setas vermelhas. E a localização da região intrônica (Figura 31) está evidenciada pelas setas rosa.

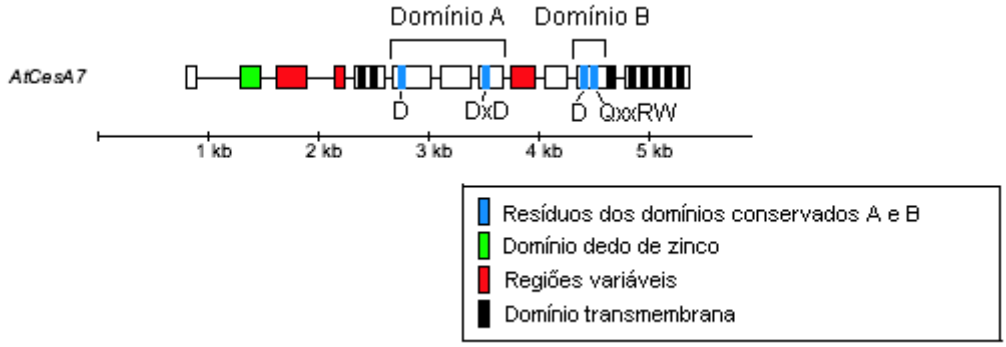


Figura 32. Estrutura do gene *AtCesA7*. Os éxons são representados pelos retângulos e os íntron pelas linhas. Adaptado de Richmond (2000).

```
CesA3  664   GMDDDKELLMSEMNFEKKFGQSAIFVTSTLMEQGGVPPSSSPAALLKEAIHVISCGYEDK  723
             GMDDDKELLMSEMNFEKKFGQSAIFVTSTLMEQGGVPPSSSPAALLKEAIHVISCGYEDK
Contig 1     GMDDDKELLMSEMNFEKKFGQSAIFVTSTLMEQGGVPPSSSPAALLKEAIHVISCGYEDK  60

CesA3  724   TEWGTELGWIYGSITEDILTGFKMHCRGWRSIYCMPKRPAFKGSAPINLSDRLNQVLRWA  783
             TEWGTE   IYGSITEDILTG KMHCRG RSIYCMPKRPAFKGSAPINLSDRLNQVLRWA
Contig 61    TEWGTE---IYGSITEDILTGSKMHCRGRRSIYCMPKRPAFKGSAPINLSDRLNQVLRWA  117

CesA3  784   LGSVEIFFSHHSPVWYGYKGGKLKWPERFAYVNTTIYPFTSLPLLAYCTLPAICLLTDKF  843
             LGSVEIFFSHHSPVWYGYKGGKLKW ERFAYVNTTIYPFTSLPLLAYCTLPAICLLTDKF
Contig 118   LGSVEIFFSHHSPVWYGYKGGKLKWLERFAYVNTTIYPFTSLPLLAYCTLPAICLLTDKF  177
                                                        DT3
CesA3  844   IMPAISTFASLFFIALFMSIFATGILELRWSGVSIEEWWRNEQFWVIGGVSAHLFAVVQG  903
             IMPAISTFASLFFIALFMSIFATGILELRWSGVSIEEWWRNEQFWVIGGVSAHLFAVVQG
Contig 178   IMPAISTFASLFFIALFMSIFATGILELRWSGVSIEEWWRNEQFWVIGGVSAHLFAVVQG  237
                         DT4                                DT5
CesA3  904   LLKVLAGIDTNFTVTSKASDDEDFGELYAFKWXXXXXXXXXXXXXXXVGVVAGISDAINN  963
             LLKVLAGIDTNFTVTSKASDDEDFGELYAFKWTTLLIPPTTILIINLVGVVAGISDAINN
Contig 238   LLKVLAGIDTNFTVTSKASDDEDFGELYAFKWTTLLIPPTTILIINLVGVVAGISDAINN  297
                                                        DT6
CesA3  964   GYQAWGPLFGKLFFAFWVILHLYPFLKGLMGRQNRTPTIVVIWSVLLASIFSLLWVRIDP  1023
             GYQAWGPLFGKLFFAFWVILHLYPFLKGLMGRQNRTPTIVVIWSVLLASIFSLLWVRIDP
Contig 298   GYQAWGPLFGKLFFAFWVILHLYPFLKGLMGRQNRTPTIVVIWSVLLASIFSLLWVRIDP  357
                         DT7                                DT8
CesA3 1024   FVLKTKGPDTKKCGI  1038
             FVLKTKGPDTKKCGI
Contig 358   FVLKTKGPDTKKCGI  372
```

Figura 33. Alinhamento entre a seqüência *CesA* de *Eucalyptus grandis* e o contig formado pelo seqüênciamento de parte do gene *CesA7.2* de *Eucalyptus* realizado pelo programa Blast.

Outro fato que pôde ser observado a cerca da estrutura do gene *CesA7.2* de *Eucalyptus* é que este possui apenas um íntron entre a região variável e os domínios trans-membrana carboxi-terminais, diferentemente do seu homólogo *AtCesA7* (Figura 32) que possui 3 íntrons entre as respectivas regiões.

5.12 ANÁLISE DOS NÍVEIS DE EXPRESSÃO DO GENE *CesA7.2* DE CELULOSE SINTASE DE *Eucalyptus*

A duas seqüências genômicas correspondentes aos *contigs* 1 e 2 também foram submetidas a uma análise *blast* contra o banco de dados de ESTs do Projeto Genolyptus para avaliar a proporção de *reads* que são similares a essas seqüências. Os dois *contigs* juntos são similares (com E-valor menor do que 10^{-7}) a 112 *reads* submetidos ao Projeto Genolyptus, distribuídos em diferentes bibliotecas construídas para espécies e tecidos distintos (Tabela 13).

Estudos de expressão diferencial são ferramentas importantes para o entendimento da ocorrência de processos biológicos ao nível molecular. Tais estudos contribuem de maneira significativa para a caracterização e identificação de redes de genes de importância econômica, que atuam no desenvolvimento de determinada estrutura ou na resposta de um organismo a um estímulo externo. Estes estudos possibilitam também a identificação de genes-chave pela comparação da expressão gênica durante o desenvolvimento de um tecido ou tipo celular em condições normais e quando submetidos a algum tipo de *stress*. Dados de expressão diferencial podem ser usados para reconstruir as vias reguladas por esses genes, predizer onde eles atuam e identificar novos genes associados ao processo de interesse (Guimarães *et al.,* 2005).

A Figura 34 representa a porcentagem de seqüências similares à *CesA7.2* expressas em diferentes espécies de *Eucalyptus*. Existe uma diferença significativa no nível de expressão destes genes no xilema de *E. urophylla* em relação às demais ($p<0,05$), embora não tenham sido detectados níveis distintos de expressão destes genes no xilema das demais espécies analisadas ($p>0,05$). Estudos posteriores, no entanto, devem ser realizados para se estabelecer a possível relação causal entre os níveis de expressão dos genes de celulose sintase e as diferenças observadas nas qualidades físico-químicas da madeira de diferentes espécies de *Eucalyptus* e *E. urophylla*. A madeira de *E. grandis*, por exemplo, é considerada leve, com uma densidade em torno de 0,700 a 0,800 g/cm^3,

enquanto a densidade da madeira em *E. pellita* está em torno de 0,960 g/cm^3. Já a madeira de *E. urophylla* é considerada medianamente leve (Revista Madeira, 2003).

Tabela 13. Distribuição de *reads* com elevado nível de similaridade aos *contigs* obtidos para o gene *CesA7.2* de *Eucalyptus grandis*, entre as bibliotecas do Projeto Genolyptus.

Nome da Biblioteca	Espécie	Tecido	Tamanho da Biblioteca	Nº de *reads* homólogos
EUGL-XY	*E. globulus*	Xilema	15703	32
EUGR-BC	*E. grandis*	BAC	17323	2
EUGR-ML	*E. grandis*	Folhas maduras	5.934	1
EUGR-PU	*E. grandis*	Folhas infectadas com puccina	8.075	1
EUGR-SE	*E. grandis*	Plântula	10.690	1
EUGR-XY	*E. grandis*	Xilema	8.463	17
EUPE-XY	*E. pellita*	Xilema	9.755	20
EUSP-FX	*E. spp.*	Floema (diversas espécies)	12.248	5
EUUR-XY	*E. urophulla*	Xilema	9.500	33

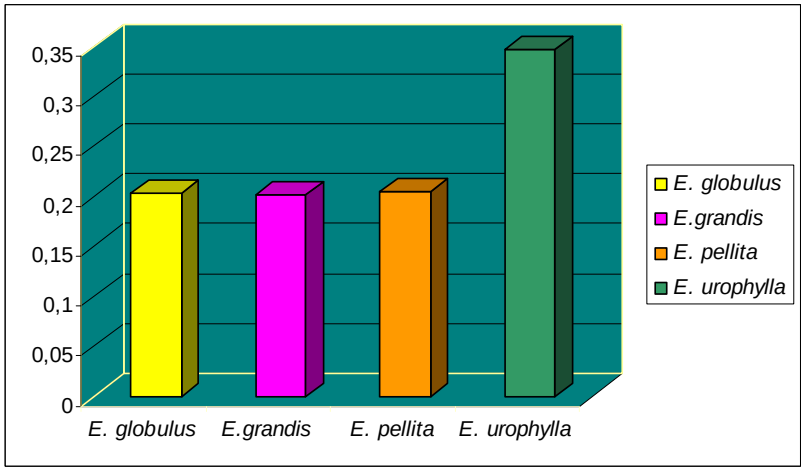


Figura 34. Porcentagem de *reads* detectados com elevado nível de similaridade ao gene *CesA7.2* em bibliotecas obtidas de xilema de diferentes espécies de *Eucalyptus*.

Como esperado, os resultados apresentados na Figura 35 sugerem que o gene *CesA7.2* seja mais expresso em xilema que em tecidos foliares ($p<0,001$). Esses resultados são semelhantes àqueles obtidos em estudos de expressão diferencial realizados recentemente em *A. thaliana*, que demonstram que a formação da parede celular secundária aumenta dramaticamente do topo para a base do caule, sendo muito reduzida nas folhas (Brown *et al.*, 2005). Os baixos níveis de expressão destes genes em folhas são consistentes com a presença de uma pequena porção de sistemas vasculares enriquecidos com celulose nesses tecidos (Suzuki *et al*., 2006).

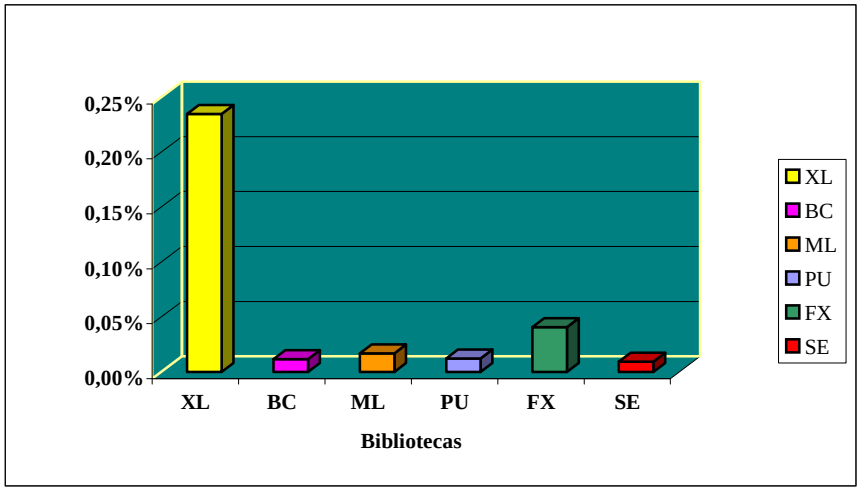


Figura 35. Porcentagem de *reads* detectados com elevado nível de similaridade ao gene *CesA7.2* em bibliotecas de diferentes tecidos em diferentes espécies de *Eucalyptus*: XL (xilema); BC (DNA genômico); ML (folhas maduras); PU (folhas infectadas com *Puccinia*); FX (biblioteca construída com floema de diferentes espécies) e SE (plântulas).

5.13 DETECÇÃO DE POLIMORFISMO NO GENE *CesA7.2* DE CELULOSE SINTASE EM DIFERENTES ESPÉCIES DO GÊNERO *Eucalyptus*

O alinhamento formado pelas seqüências de DNA genômico de diferentes espécies de *Eucalyptus* obtido nos programas *Staden Package* e *ClustalX* (Figura 36) compreende um fragmento de 501 pares de bases. A seqüência deste fragmento foi comparada à seqüência de cDNA de *Eucalyptus* depositada no GenBank como *CesA3*. O programa EST2GENOME foi utilizado para a localização do íntron (Figura 37). O alinhamento realizado entre o *contig* e a seqüência de cDNA de *Eucalyptus* confirmou a localização da região intrônica apresentada na Figura 37. Esta é formada por 95 pares de bases e está localizada entre as bases 323 e 417 do *contig*. Na figura 37 é possível observar duas mutações que ocorre em regiões exônicas. A primeira mutação ocorre na primeira base de um códon, uma alteração de triptofano (aminoácido apolar, com cadeia lateral hidrofóbica), observado na seqüência CesA3 de *Eucalyptus grandis*, para uma arginina (aminoácido básico, com cadeia lateral hidrofílica) no contig. A substituição que ocorre nesse ponto causa uma mutação do tipo não-sinônima. A segunda mutação é do tipo

sinônima, já que o aminoácido prolina que ocorre em CesA3 de *Eucalyptus grandis* possui as mesmas caracerísticas bioquímicas do amínoácido leucina que é observado no contig.

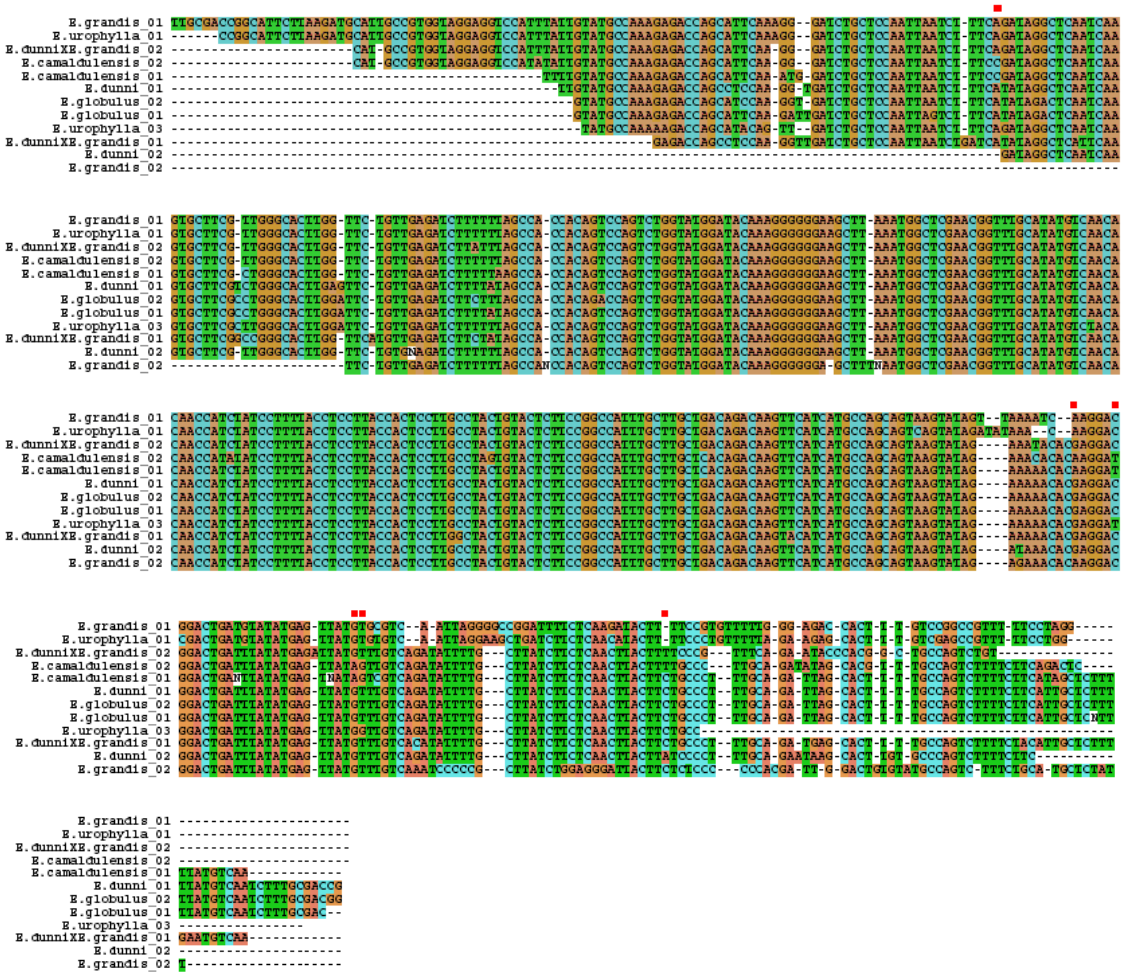


Figura 36. Alinhamento de seqüências genômicas de um fragmento do gene *CesA7.2* de diferentes espécies de *Eucalyptus* obtido pelo programa CLUSTAX. Os quadrados vermelhos indicam a localização dos SNPs.

```
Contig1      16 ttaagatgcattgccgtggtaggaggtccatttattgtatgccaaagaga     65
                ||||||||||||||||||||| |||||||||||||||||||||||||||||||
  CesA3    2267 TTAAGATGCATTGCCGTGGTTGGAGGTCCATTTATTGTATGCCAAAGAGA   2316

Contig1      66 ccagcattcaagggatctgctccaattaatctttcagataggctcaatca    115
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
  CesA3    2317 CCAGCATTCAAGGGATCTGCTCCAATTAATCTTTCAGATAGGCTCAATCA   2366

Contig1     116 agtgcttcgttgggcacttggttctgttgagatctttttagccaccaca    165
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
  CesA3    2367 AGTGCTTCGTTGGGCACTTGGTTCTGTTGAGATCTTTTTTAGCCACCACA   2416

Contig1     166 gtccagtctggtatggatacaaagggggggaagcttaaatggctcgaacgg    215
                ||||||||||||||||||||||||||||||||||||||||||| |||||||
  CesA3    2417 GTCCAGTCTGGTATGGATACAAAGGGGGGAAGCTTAAATGGCCCGAACGG   2466
```

```
Contig1    216 tttgcatatgtcaacacaaccatctatccttttacctccttaccactcct    265
               ||||||||||||||||||||||||||||||||||||||||||||||||||
  CesA3   2467 TTTGCATATGTCAACACAACCATCTATCCTTTTACCTCCTTACCACTCCT   2516

Contig1    266 tgcctactgtactcttccggccatttgcttgctgacagacaagttcatca    315
               ||||||||||||||||||||||||||||||||||||||||||||||||||
  CesA3   2517 TGCCTACTGTACTCTTCCGGCCATTTGCTTGCTGACAGACAAGTTCATCA   2566

Contig1    316 tgccagcagtaag....tgcagattagcacttttgccagtcttttcttca    446
               ||||||||>>>>> 95 >>>>>|||||||||||||||||||||||||||||
  CesA3   2567 TGCCAGCA.............ATTAGCACTTTTGCCAGTCTTTTCTTCA   2602

Contig1    447 ttgctctttttatgtcaatctttgcgaccg    476
               |||||||||||||||||||||||| ||||
  CesA3   2603 TTGCTCTTTTTATGTCAATCTTTGCAACCG   2632
```

Figura 37. Alinhamento obtido pelo programa EST2GENOME evidenciando a localização do íntron. As setas indicam bases divergentes entre a seqüência do contig e a seqüência CesA3 de *Eucalyptus grandis*.

A diversidade nucleotídica (π) foi estimada nas regiões exônicas e intrônicas, bem como em todo fragmento utilizando os programas *MEGA3.1* (Kumar *et al*., 2004) e *DnaSP* (Rozas & Rozas, 1999) e estes valores foram coincidentes em ambos (Tabela 14). O erro padrão da diversidade nucleotídica foi calculado com 10000 ciclos de *bootstrap* pelo programa *MEGA 3.1* e o número de sítios polimórficos foi calculado pelo programa *DnaSP* (Tabela 15).

Tabela 14. Caracterização do polimorfismo detectado em um fragmento do gene *CesA7.2* em diferentes espécies do gênero *Eucalyptus*.

Parâmetros	Exon 1	Intron	Exon 2	Total
Nº de sítios	315	88	77	480
S (nº de sítios polimorficos)	18	26	31	75
Diversidade nucleotídica (π)	0,0201	0,0824	0,2029	0,0543
Erro Padrão	0,0053	0,0142	0,0272	0,0061

O nível de diversidade nucleotídica observado no segundo éxon foi muito maior que a diversidade encontrada nas outras regiões onde este parâmetro foi estimado. À semalhança de outros trabalhos realizados com outros genes, a diversidade encontrada na região intrônica superou aquela encontrada no éxon 1 (Santos, 2005; García-Gil *et al*., 2003). Curiosamente, o nível de polimorfismo encontrado no éxon 2 foi bastante elevado, embora esta variação esteja concentrada em uma região de aproximadamente 30 pares de bases. De modo geral, a diversidade nucleotídica total encontrada para esse fragmento em *Eucalyptus* (0,0543) foi maior que a diversidade encontrada em um gene relacionado com

a produção de parede celular secundária de *Pinus taeda*, cuja diversidade encontrada foi de π=0,00087 para um fragmento de aproximadamente 1020 pares de bases (Brown *et al*., 2004). Esse gene de *Pinus taeda* possui 82% de identidade com o gene *CesA3* de *Eucalyptus*, que por sua vez é homólogo à *AtCesA7*.

A proporção de sítios polimórficos também foi estimada dentro das três regiões (Tabela 15), evidenciando que o maior polimorfismo está preservado dentro do segundo éxon.

Tabela 15. Proporção de sítios polimórficos em regiões intrônicas e exônicas do gene *CesA7.2* em diferentes espécies de *Eucalyptus*.

Parâmetro	Exon 1	Intron	Exon 2
Proporção de sítios polimórficos	0,0571	0,2954	0,4026

Apesar da grande quantidade de sítios polimórficos encontrada no fragmento, uma análise mais rigorosa para detecção de SNPs reais nessa região foi realizada com o auxílio do programa *StadenPackage*, utilizando os parâmetros descritos na Figura 38. Foram detectados, por este programa, somente 6 SNPs na região flanqueada pelos *primers* CesA72.5 e CesA72.6. Os SNPs foram nomeados pela sua posição no *contig* à partir da base número 1 que forma essa região (Figura 38).

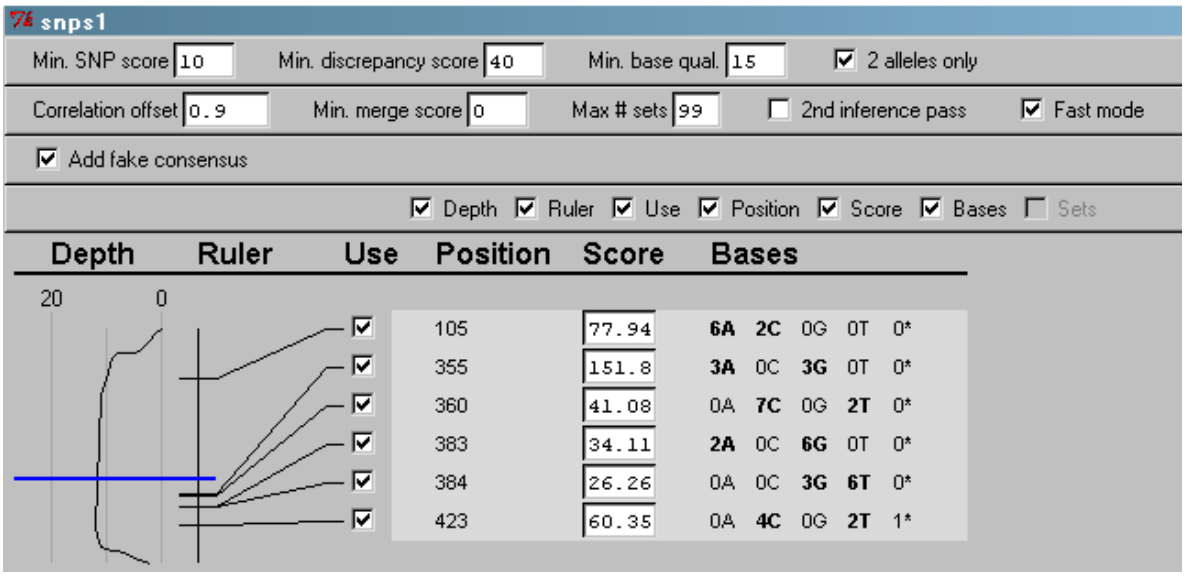


Figura 38. Localização dos 6 SNPs encontrados pelo programa *Staden Package* na região flanqueada pelos *primers* CesA72.5 e CesA72.6.

Dentre os seis SNPs encontrados, em quatro deles ocorreram transições, ou seja a substituição de uma purina por uma purina ou uma pirimidina por uma pirimidina. E em dois ocorreram transversões, ou seja, substituição de purinas por pirimidinas e vice-versa (Tabela 16). De modo geral, em seqüências de DNA a taxa de transição é superior à taxa de transversão. Esse fato é condizente com o observado para as mutações relativas ao fragmento do gene de celulose sintase seqüenciado.

Tabela 16. Caracterização dos SNPs detectados em relação ao tipo de substituição.

Posição do SNP	Tipo de substituição	Bases
105	Transversão	A/C
355	Transição	G/A
360	Transição	C/T
383	Transição	A/G
384	Transversão	G/T
423	Transição	C/T

Figura 39. Localização dos 6 SNPs encontrados no fragmento seqüenciado pelos *primers* CesA72.5 e CesA72.6.

Somente dois desses SNPs estão localizados em regiões exônicas (Figura 39). O SNP da posição 105 está localizado no primeiro éxon do fragmento e é uma tranversão que ocorre entre os nucleotídeos A e C. O nucleotídeo C aparece exclusivamente em uma das espécies (*Eucalyptus camaldulensis*) (Figura 40). De acordo com a ORF encontrada para essa seqüência no programa GeneRunner, a tranversão que ocorre na 3º base do códon causa uma mutação silenciosa (Griffiths *et al*., 2002), alterando um códon TCA (Serina) para um códon TCC (Serina), fato que é esperado para a maioria das substituições que ocorrem nessa posição do códon (Nei & Kumar, 2000). Desse modo, o SNP não altera a seqüência de aminoácidos da proteína nesse ponto, embora alterações no nível de expressão possam ocorrer pela utilização preferencial de determinados códons de serina.

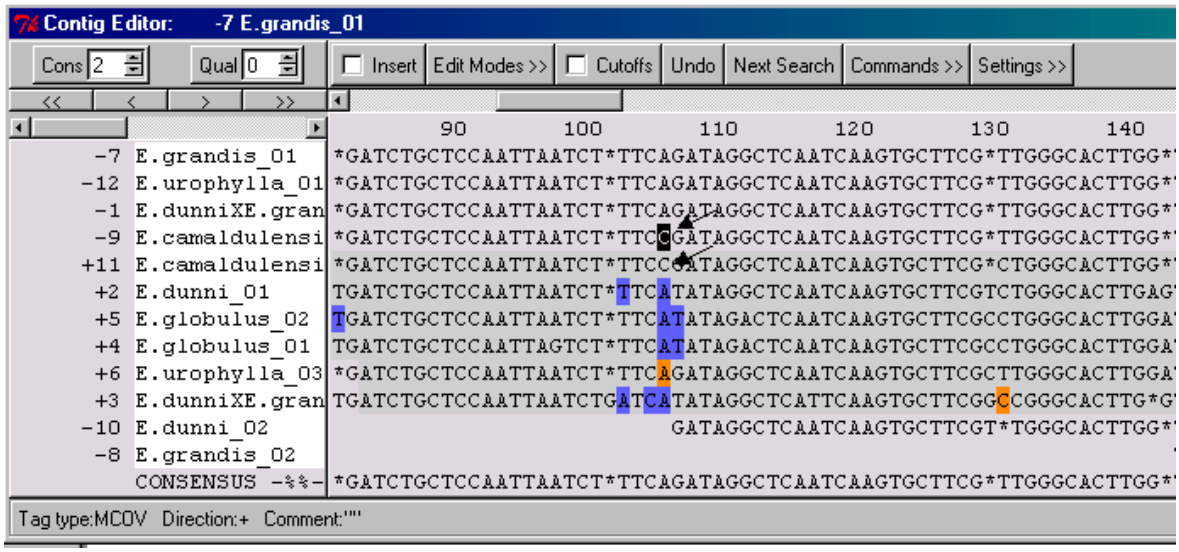


Figura 40. Alinhamento obtido pelo programa *Staden Package* evidenciando a transversão que ocorre em apenas uma das espécies: *E. camaldulensis* (C1).

A inspeção visual desse SNP foi um grande suporte para validar a existência do mesmo e este pode ser visualizado na Figura 41. Os dois primeiros quadros são relativos à uma região do eletroferograma de duas espécies de *Eucayptus* diferentes (*E. camaldulensis* e o híbrido *E. dunni X E. grandis*, respectivamente) abrangendo o SNP da posição 105 (seta vermelha). O terceiro quadro exibe as diferenças entre as regiões realizadas pelo módulo *Trace Difference* do programa *Staden Package.*

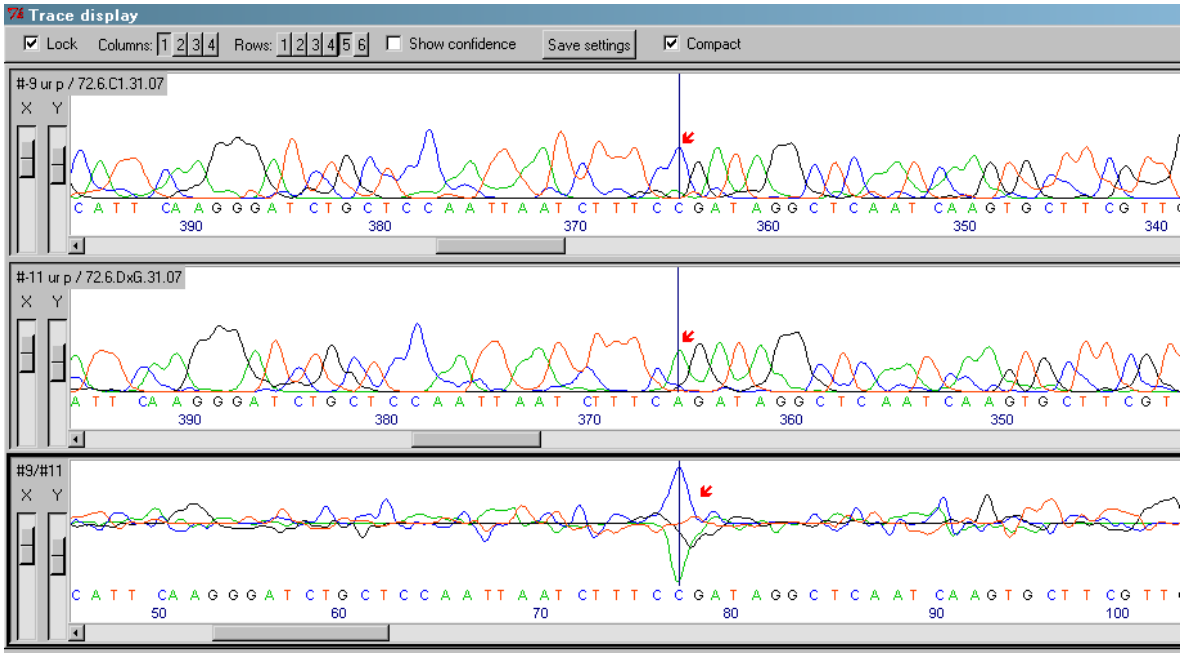


Figura 41. Eletroferogramas representativos da região abrangendo SNP da posição 105.

O SNP na posição 423 também está localizado em região exônica (no segundo éxon do fragmento), este é uma transição de C para T na segunda base do códon e corresponde a uma mutação não-silenciosa causando uma alteração no aminoácido codificado por esse códon. Todas as substituições nucleotídicas que ocorrem na segunda posição do códon geram mutações desse tipo (Nei & Kumar, 2000). A alteração que ocorre é de uma serina (TCT) que é um aminoácido polar neutro para uma fenilalanina (TTT) que é um aminoácido com característica apolar. Tal mudança, não-sinônima, pode alterar, além da seqüência de aminoácidos, a conformação estrutural que esta proteína pode assumir. Esse SNP ocorre em indivíduos de espécies diferentes de *Eucalyptus*, como indicado pelas setas vermelhas da Figura 42 mostrando o híbrido *E. dunni* X *E. grandis* e um indivíduo de *E. urophylla*, bem como no mesmo indivíduo de *E. camaldulensis* (setas azuis da Figura 42) indicando a possível existência de heterozigose nesse ponto. A inspeção visual, neste caso, não confirma a existência de um SNP nesse ponto.

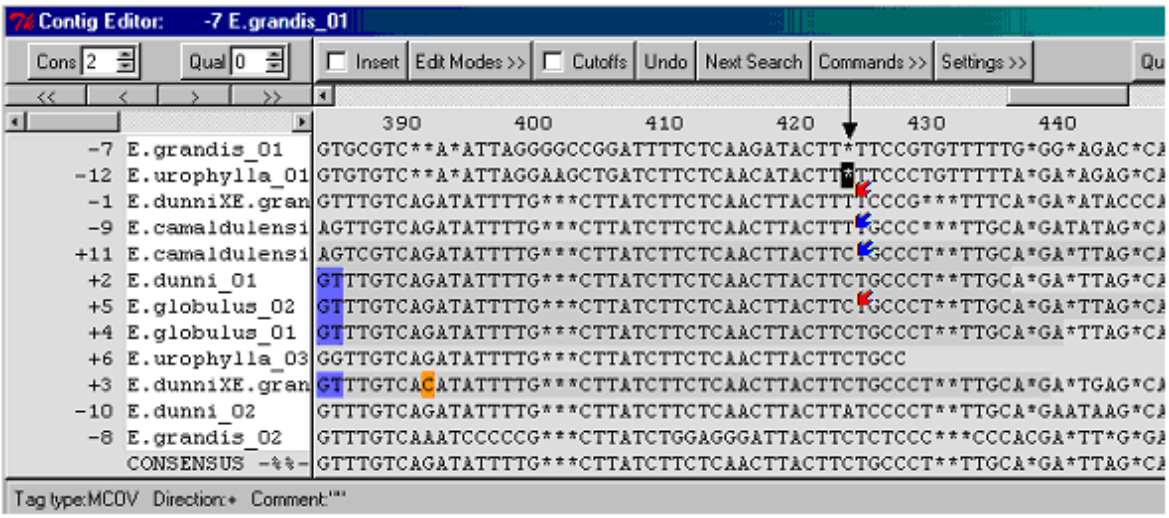


Figura 42. Alinhamento evidenciando a presença de um SNP na posição 423 do segundo éxon.

Os outros quatro SNPs (posições: 355, 360, 383 e 384) estão localizados dentro do íntron (Figura 43). Diferentemente das estimativas de diversidade nucleotídicas mostradas na Tabela 13 a maior variação representada por SNPs está situada dentro da região intrônica. Esse fato é plausível já que a pressão de seleção exercida sobre esta região é menor (Ching *et al.*, 2002).

A maior variação encontrada está entre espécies, sendo *E. camaldulensis* o mais divergente de acordo com os SNPs identificados. A divergência contida nos SNPs neste caso, no entanto, não é transmitida para a proteína, já que o SNP na posição 105 é

responsável por uma mutação silenciosa e os outros quatro SNPs estão representados na regiões intrônica.

Figura 43. Alinhamento evidenciando a ocorrência de prováveis SNPs na região intrônica.

6 CONSIDERAÇÕES FINAIS

Os resultados obtidos por este trabalho devem ser considerados como uma contribuição preliminar para a realização de estudos posteriores de caracterização da estrutura e dos níveis de polimorfismo presentes em genes de celulose sintase de parede secundária em espécies do gênero *Eucalyptus.* A partir da anotação dos ESTs obtidos nas bibliotecas de cDNA do projeto Genolyptus, pela utilização de ferramentas de análise filogenética, foram construídos alinhamentos a partir dos quais foram desenhados *primers* específicos para cada um dos genes de interesse, perfazendo um total de 28 *primers*. A disponibilização destes *primers* constitue uma etapa primordial para o desenvolvimento dos estudos de caracterização e manipulação destes genes.

Um exemplo de aplicação foi produzido pela utilização destes *primers* na triagem da biblioteca de BACs de *Eucalyptus grandis* a fim de se realizar a caracterização do gene *CesA7.2* sob diferentes aspectos. Embora a seqüência de 1396pb obtida para este gene não seja inédita, cumpre ressaltar que o esforço aqui realizado se deu de maneira independente, e ainda que a seqüência relativa ao íntron de 92pb é inédita. A partir desse fragmento seqüenciado, estudos de níveis de expressão desse gene também foram realizados para diferentes espécies de *Eucalyptus* demonstrando uma maior expressão desse gene em xilema que em tecidos foliares, confirmando a atuação desse gene na formação da parede celular secundária. Além disso, o nível de expressão desse gene foi maior em *E. urophylla* ($p < 0,05$) que em outras espécies analisadas ($p > 0,05$).

Destaca-se ainda a importância dos resultados aqui obtidos em termos de detecção dos sítios polimórficos distribuídos ao longo do fragmento seqüenciado do gene *CesA7.2* em diferentes espécies de *Eucalyptus.* O índice de diversidade nucleotídica foi calculado para os dois éxons do fragmento bem como para o íntron, sendo esse índice superior no segundo éxon. Os seis SNPs detectados, dois em regiões exônicas e quatro em regiões intrônicas, podem ser utilizados em estudos posteriores objetivando o mapeamento genético do gene *CesA7.2*, para fins de aplicação da estratégia conhecida como de genes candidatos. A aplicação desta estratégia permitiria se avaliar a hipótese de que o

polimorfismo aí presente possa ser responsável pela variação fenotípica observada em características físico-químicas da madeira de *Eucalyptus* relacionadas com a qualidade e quantidade de celulose. Os SNPs aqui detectados podem ainda ser utilizados para se avaliar a extensão do desequilíbrio de ligação presente no contexto deste gene, buscando se avaliar o potencial de utilização das estratégias conhecidas como mapeamento por associação.

Por fim, considera-se que os métodos aqui utilizados representam uma iniciativa no sentido de se utilizar as ferramentas de análise genômica para a caracterização de genes, putativamente cruciais para o melhoramento genético de características físico-químicas da madeira de *Eucalyptus*, principalmente no que tange à quantidade e qualidade da celulose produzida para fins industriais.

7 REFERÊNCIAS BIBLIOGRÁFICAS

AUDIC, S. & CLAVERIE, J. M. The significance of digital gene expression profiles.**Genome Research**. v. 7, n.10, p. 986-995.1997.

ALTSCHUL S. F.; GISH, W.; MILLER, W. MYERS, E. W. & LIPMAN, D. J. Basic Local Alignment Search Tool. **Journal of Molecular Biology.** v. 215, p. 403-410.1990.

ANDERSSON-GUNNERAS, S.; MELLEROWICZ, E. J.; LOVE, J.; SEGERMAN, B.; OHMIYA, Y.; COUTINHO, P. M.; NILSSON, P.; HENRISSAT, B.; MORITZ, T. & SUNDBERG, B. Biosynthesis of cellulose-enriched tension wood in *Populus:* global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. **The Plant Journal.** v. 45, p. 144-165.2006.

BERTOLA, A. **"Eucalipto: Verdades e Mentiras".** Disponível em: www.celuloseonline.com.br/imagembank/Docs/DocBank/dc/dc009.pdf. Acesso em 20 de julho de 2006.

BONFIELD, J. K.; RADA, C. & STADEN, R. Automated detection of point mutations using fluorescent sequence trace subtraction. **Nucleic Acids Research**. v. 26, p. 3404-3409. 1998.

BRACELPA (a) – Associação Brasileira de Celulose e Papel. **Florestas Plantadas com Eucalipto e Pinus – Ecologia das florestas plantadas.** Disponível em:

http://www.bracelpa.org.br/br/estudantes/floresta/ecologia.htm. Acesso em 15 junho de 2006.

BRACELPA (b) – Associação Brasileira de Celulose e Papel. **O Setor Brasileiro de Papel e Celulose.** Disponível em: http://www.bracelpa.org.br/br/news/anteriores.htm. Acesso em 15 junho de 2006.

BROOKES, A. J. The Essence of SNPs. **Gene.** v. 234, p. 177-186. 1999.

BROWN Jr., R. M. & SAXENA, I. M. Cellulose biosynthesis: A moldel for understanding the assembly complex biopolymers. **Plant Physiology Biochemistry.** v. 38, n. 1-2, p. 57-67.2000.

BROWN, G. R.; GILL, G. P.; KUNTZ, R. J.; LANGLEY, C. H. & NEALE, D. Nucleotide diversity and linkage disequilibrium in loblolly pine. **Proceedings of the National Academy of Science of the Unites States of America - PNAS.** v. 19, n. 42, p. 15255-15260. 2004.

BURTON, R. A.; FARROKHI, N.; BACIC, A. & FINCHER, G. B. Plant cell wall polysaccharide biosynthesis: real progress in the identification of participating genes. **Planta.** v. 221, p. 309-312. 2005.

CHAKRAVARTI, A. Silgle Nucleotide Polymorphisms- ...to a future of genetic medicine. **Nature.** v. 409, p. 822-823. 2001.

DELMER, D. P. & AMOR, Y. Cellulose Biosynthesis. **Thel Plant Cell.** v. 7, p. 987-1000.1995.

DHUGGA, K. S. Building the wall: genes and enzyme complexes for polysaccharide synthases. **Current Opinion in Plant Biology.** v. 4, p 488-493. 2001.

DICKINSON, J. C. The Eucalypt in the Sierra of Southern Peru. **Annals of the Association of American Geographers.** v. 59, n. 2, p. 294-306.1969.

DOBLIN, M. S.; KUREK, I.; JACOB-WILK, D. & DELMER, D. P. Cellulose biosynthesis in Plants: from genes to rosettes. **Plant Cell Physiology.** v. 43, n.12, p. 1407-1420. 2002.

EUCLID – Eucalypts of Southern Austrália. Disponível em: www.anbg.gov.au/cpbr/cd-keys/**Euclid**/index.html. Acesso em 14 de agosto de 2006.

EWING, B.; HILLIER, L.; WENDL, M. C. & GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy Assessment. **Genome Research.** v. 8, p. 175-185. 1998.

FAGARD, M.; DESNOS, T.; DESPREZ, T.; GOUBET, F.; REFREGIER, G.; MOUILLE, G.; MCCANN, M.; RAYON, C.; VERNHETTES, S. & HOFTE, H. *PROCUSTE 1* Encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotys of *Arabidopsis.* **The Plant Cell.** v. 12, p. 2409-2423. 2002.

FAO – FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Disponível em: http://www.fao.org/forestry. Acesso em 12 de setembro de 2006.

FELSENSTEIN, J. **Inferring Phylogenies.** Massachusetts: Sinauer Associates. 2004. 580p.

FERREIRA, M. Escolha de Espécies de Eucalipto. **Circular Técnica IPEF.** v. 47, p. 1-30.1979.

FERREIRA, M. E. & GRATTAPAGLI, D. **Introdução ao uso de marcadores moleculares em análise genética.** 3º Ed. Brasília: Embrapa-Cenargen, 1998. 220p.

GARDINER, J.C.; TAYLOR, N.G.; TURNER, S.R. Control of Cellulose Complex Localization in Developing Xylem. **The Plant Cell.** v. 15, p. 1740 – 1748. 2003.

GIBAS, C. & JAMBECK, P. **Desenvolvendo bioinformática.** Rio de Janeiro: Campus, 2001.464p.

GONZÁLEZ, E. R.; ANDRADE, A.; BERTOLO, A. L.; CARNEIRO, R. T.; LACERDA, G. C.; DEFÁVARI, V. A. P.; LABATE, M. T. V. & LABATE, C. A. Transformação genética do Eucalipto. **Biotecnologia, Ciência e Desenvolvimento.** v. 26, p. 18-22. 2002.

GRIFFITHS, A. J. F.; MILLER, J. H.; SUZUKI, D. T.; LEWONTIN, R. C. & GELBART, W. M. **Introdução à genética.** 6º ed. Rio de Janeiro: Guanabara Koogan, 2002. 794p.

GUIMARÃES, P. E. M. & COSTA, M. C. R. SNPs: Sutis diferenças de um código. **Biotecnologia, Ciência e Desenvolvimento.** v. 26, p. 24-27. 2002.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series** v. 41, p. 95-98. 1999.

HAMANN, T.; OSBORNE, E.; YOUNGS, H. L.; MISSON, J.; NUSSAUME, L. & SOMERVILLE, C. Global expression analysis of *CESA* AND *CSL* genes in Arabidopsis. **Cellulose.** v.11, p. 279-286. 2004.

HOLLAND, N.; HOLLAND, D.; HELENTJARIS, T.; DHUGGA, K. S.; XOCONOSTLE-CAZARES, B. & DELMER, D. P. A comparative analysis of the plant cellulose synthase (*CesA)* gene family. **Plant Physiology.** v. 123, p. 1313-1323.2002.

HOLLAND, N.; HOLLAND, D.; HELENTJARIS, T.; DHUGGA, K. S.; XOCONOSTLE-CAZARES, B. & DELMER, D. P. A comparative analysis of the plant cellulose synthase (CesA) gene family. **Plant Physiology.** v. 123, n. 4, p. 1313-1324. 2000.

IPEF – Instituto de Pesquisas e Estudos Florestais. **Chave de Identificação de Espécies Florestais.** Disponível em: http://www.ipef.br/identificacao/cief/especies/grandis.asp. Acesso em 12 de maio de 2006.

JEANMOUGIN, F.; THOMPSON, J. D.; GOUY, M.; HIGGINS, D. G. & GIBSON, T. J. Multiple sequence alignment with Clustal X**. Trends Biochemical Sciences**. v. 23, n. 10, p.403-405. 1998.

JOSHI, C. P.; BHANDARI, S.; RANJAN, P.; KALLURI, U. C.; LIANG, X.; FUJINO, T. & SAMUGA, A. Genomics of cellulose biosynthesis in poplars. **New Physiologist.** v.164, p. 53-61.2004.

KALLURI, U. C. & JOSHI, C. P. Differential expression patterns of two cellulose synthase genes are associated with primary and secondary cell wall development in aspen trees. **Planta.** v. 220, n. 1, p. 47-55. 2004.

KUMAR, S.; TAMURA, K. & NEI M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. **Briefings in Bioinformatics.** v. 5, p.150-163. 2004.

KWOK, P. & GU, Z. Single nucleotide polymorphism libraries: why and how are we building them? **Molecular Medicine Today.** v.5, p. 538-543. 1999.

LEHNINGER, A. L.; NELSON, D. L. & COX, M. M. **Princípios de Bioquímica.** São Paulo: Sarvier, 1992. 839p.

LIMA, W. P. **Impacto Ambiental do Eucalipto.** São Paulo: EDUSP, 1993. 301p.

LODISH, H.; BERK, A.; ZIPURSKY, S. L.; MATSUDAIRA, P.; BALTIMORE, D. & DARNELL, J. E. **Molecular Cell Biology.** 4º ed. New York. W. H. Freeman & Co., 2000. 1084p.

MOTT, R. EST_GENOME: a program to align splices DNA sequences to unspliced genomic DNA. **Computer Applications in the Biosciences,** v. 13, n. 4, p. 477-478. 1997.

NEI, M. **Molecular Evolutionary Genetics**. New York: Columbia University Press, 1987. 512p.

PAGE, R. D. TreeView: an application to display phylogenetic trees on personal computers. **Computer Applications in the Biosciences.** v. 12, n.4, p. 357-358. 1996.

PAULA, R. N. Ferramentas e estratégias melhoram eucalipto. **Revista Madeira.** v.88, ano15. 2005.

PIGATO, S. M. P. C. & LOPES, C. R. Caracterização silvicultural, botânica e avaliação da variabilidade genética por meio do marcador molecular RAPD em um teste de progênies de *Eucalyptus urophylla* S. T. Blake. **Scientia Forestalis.** v. 60, p. 135-148. 2001.

PROJETO GENOLYPTUS – Rede Brasileira de Pesquisa do Genoma de *Eucalyptus.* Proposta submetida ao MCT em 04/05/2001.

PROSDOCINI, F.; CERQUEIRA, G. C.; BINNECK, E.; SILVA, A. F.; REIS, N. A.; JUNQUEIRA, A. C. M.; SANTOS, A. C. F.; NHANI JÚNIOR, A.; WUST, C. I.; CAMARGO FILHO, F.; KESSEDIJAN, J. L.; PETRETSKI, J. H.; CAMARGO, L. P.; FERREIRA, R. G. M.; LIMA, R. P.; PEREIRA, R. M.; JARDIM, S.; SAMPAIO, V. S. & FOLGUERAS-FLATSCHART, A. V. Bioinformática: Manual do Usuário. **Biotecnologia Ciência & desenvolvimento.** n. 29, p. 12-25. 2002.

RAVEN, P. H.; EVERT, R. F. & EICHHORN, S. E. **Biologia Vegetal.** 6. ed. Rio de Janeiro: Guanabara Koogan, 2001. 906p.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology.** v. 5, p. 94-100.2002.

RANIK, M & MYBURG, A. A. Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. **Tree Physiology.** v. 26 (5), p. 545-556. 2006.

REVISTA MADEIRA. **Projeto Genoma do Eucalipto.** v.75, ano 13. 2003.

RICHMOND, T. Higher plant cellulose synthases. **Genome Biology.** v. 1, n.4, p. 3001.1-3001.6. 2002.

RICHMOND, T. A. & SOMERVILLE, C. R. The cellulose synthase superfamily. **Plant Physiology.** v. 124, n. 2, p. 495-498. 2000.

ROZEN, S. & SKALETSKY, H. J. Primer 3. Primer 3 on WWW for general users and for biologist programmers. **Methods in Molecular Biology.** v. 132, p. 365-386. 2000.

SAEED, A. I.; SHAROV, V.; WHITE, J.; LI, J.; LIANG, W.; BHAGABATI, N.; BRAISTED, J.; KLAPA, M.; CURRIER, T.; THIAGARAJAN, M.; STURN, A.; SNUFFIN, M.; REZANTSEV, A.; POPOV, D.; RYLTSOV, A.; KOSTUKOVICH, E.; BORISOVSKY, I.; LIU, Z.; VINSAVICH, A.; TRUSH, V. & QUACKENBUSH, J. TM4: a free, open-source system for microarray data management and analysis. **Biotechniques**. v. 34, n. 2, p. 374-378. 2003.

SAITOU, N. & NEI, M. The neighbor-joining method: a new method for reconstructin phylogenetic trees. **Molecular Biology and Evolution.** v. 4, p. 405-425. 1997.

SALISBURY, F. B. & ROSS, C. W. **Plant Physiology.** 4. ed. Belmont: Wassworth Publishing Company, 1991. 682p.

SAMUGA, A. & JOSHI, C. P. A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to *Arabidopsis AtCesA7* (irx3) gene associated with secondary cell wall synthesis. **GENE.** v. 296, p. 37-44. 2002.

SAMUGA, A. & JOSHI, C. P. Differential expression patterns of two new primary cell wall-related cellulose synthase cDNAs, *PtrCesA6* and *PtrCesA7* from aspen trees. **GENE.** v. 334, p. 73-82. 2004.

SANTOS, S. N. **Genes de Lignificação em *Eucalyptus:* estrutura e diversidade genética dos genes *4cl* e *ccoaomt.*** 2005. 209p. Dissertação (mestrado) – Universidade Católica de Brasília, Brasília, 2005.

SCANAVACA JUNIOR, L. **Caracterização silvicultural, botânica e tecnológica do *Eucalyptus urophylla* S.T. Blake e de seu potencial para utilização em serraria.** 2001. 108p. Dissertação (mestrado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2002.

SCHEIBLE, W.; ESCHED, R.; RICHMOND, T.; DELMER, D. & SOMERVILLE, C. Modifications of cellulose synthase confer resistence to isoxaben and thiazolidinone herbicides in *Arabidopsis Ixr1* mutants. **Proceedings of the National Academy of Science of the Unites States of America - PNAS.** v. 98, n.18, p. 10079-10084. 2001.

STADEN, R.; BEAL, K. F. & BONFIELD, J. K. The staden package. **Methods in Molecular Biology.** v. 132, p. 115-130. 1998.

STONEKING, M. Single Nucleotide Polymorphisms- From de evolutionary past....**Nature.** v. 409, p. 821-822. 2001.

SUZUKI, S.; LI, L.; SUN, Y. & CHIANG, V. L. The cellulose synthase gene superfamily and biochemical functional of xylem-specific cellulose synthase-like genes in *Populus trichocarpa.* **Plant Physiology.** v. 142, p. 1233-1245. 2006.

TAIZ, L. & ZEIGER, E. **Plant Physiology**. 2 ed. Sunderlang: Sinauer Associates, Inc., Publishers, 1998. 792p.

TAYLOR, N. G.; GARDINER, J. C.; WHITEMAN, R. & TURNER, S. R. Cellulose synthesis in the Arabidopsis secondary cell wall. **Cellulose.** v. 11, p. 329-338.2004.

TAYLOR, N. G.; HOWELLS, R. M.; HUTTLY, A. K.; VICKERS, K. & TURNER, S. R. Interactions among three distinct CesA proteins essential for cellulose synthesis. **Proceedings of the National Academy of Science of the Unites States of America - PNAS.** v. 100, n.3, p. 1450-1455.2003.

TAYLOR, N. G.; SCHEIBLE, W. CUTLER, S.; SOMERVILLE, C. R. & TURNER, S. R. The *irregular xylem3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. **The Plant Cell.** v.11, p. 769-779.1999.

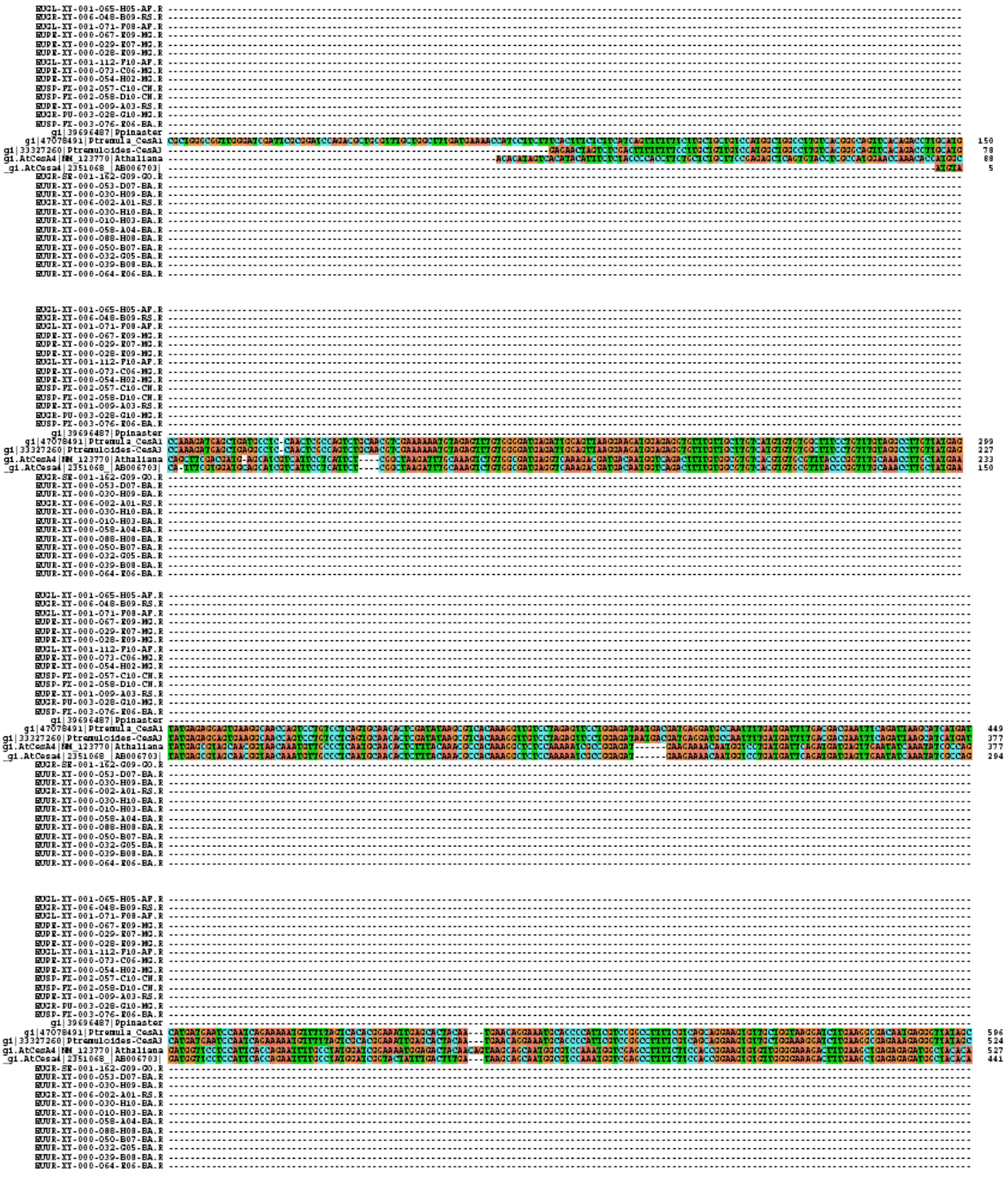
THOMPSON, J. D.; GIBSON, T. J.; PLEWNIAK, F.; JEANMOUGIN, F. & HIGGING, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleid Acid Researchs.** v. 25, n. 24, p. 4876-4882. 1997.

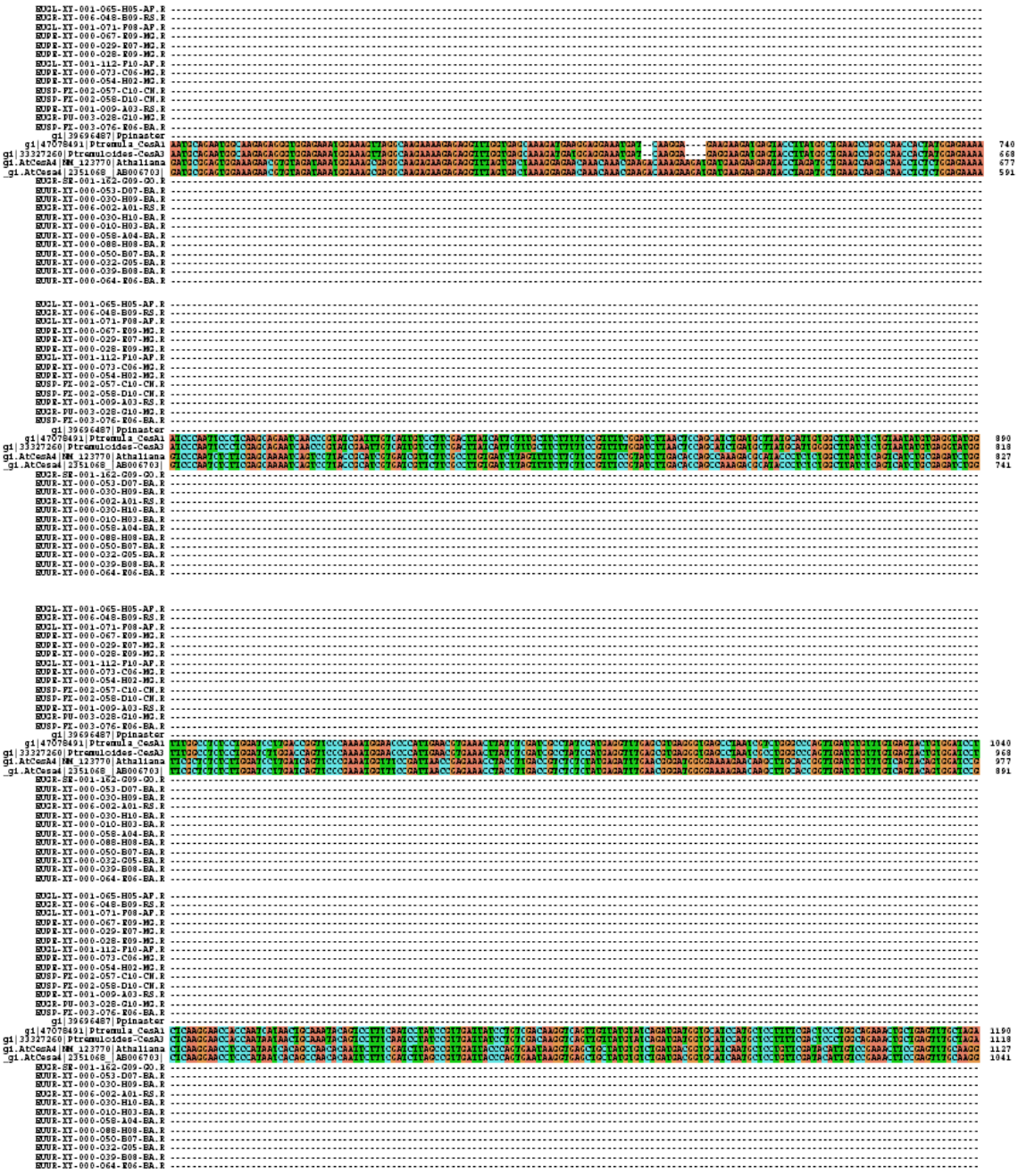
VAZ, A. R. C. **Posicionamento de microssatélites em clones BAC: rumo a integração do mapa genético e mapa físico (BACs) de Eucalyptus grandis**. 2005. 38p. Monografia -Trabalho de Conclusão de Curso (Graduação em Biologia) – Universidade Católica de Goiás, Goiânia, 2005.

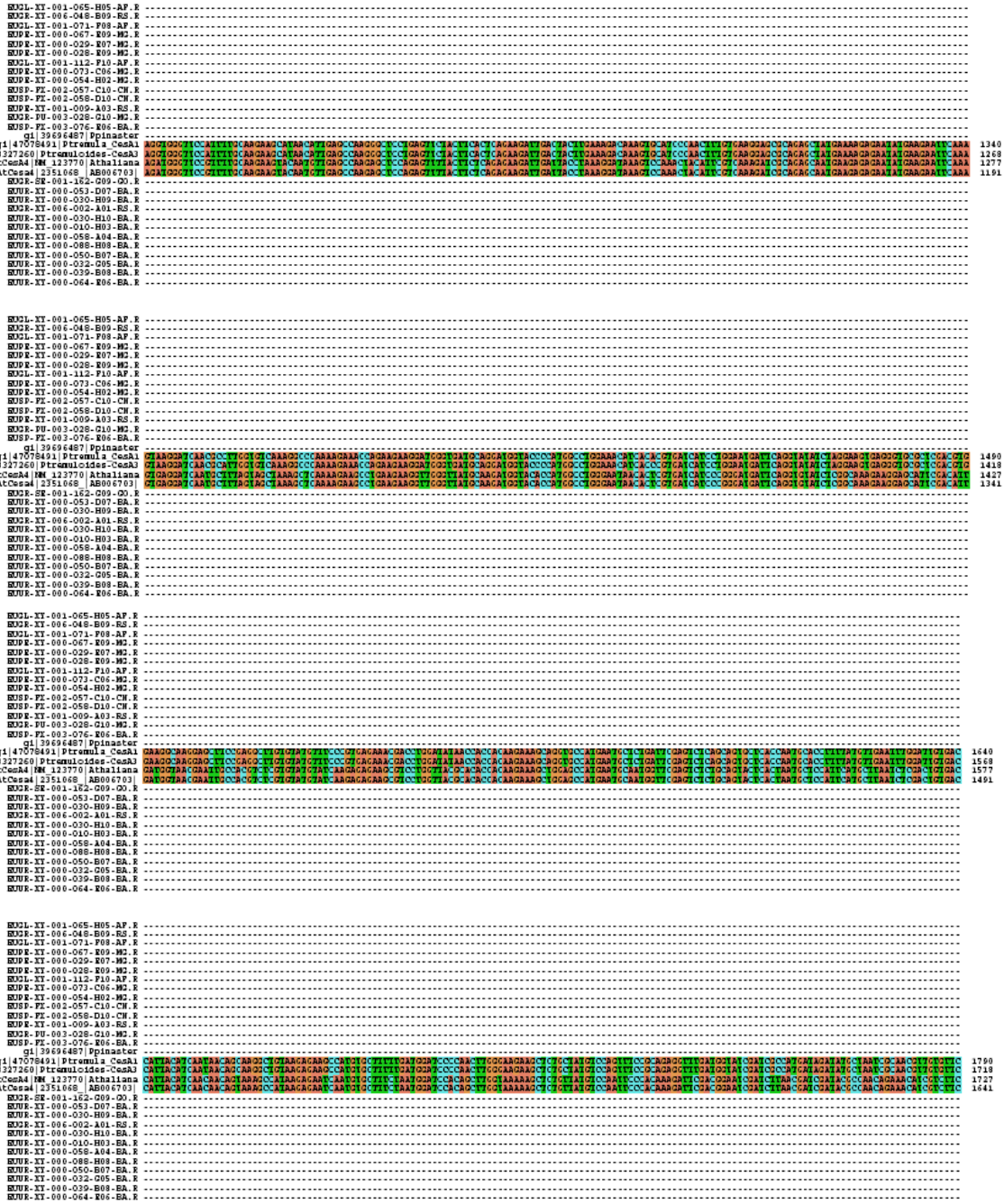
ZHONG, R.; MORRISON III, W. H.; FRESHOUR, G. D.; HAHN, M. G. & YE, Z. Expression of a mutant form of cellulose synthase *AtCesA7* causes dominant negative effect on cellulose biosynthesis. **Plant physiology.** vol. 132, p. 786-795. 2003.

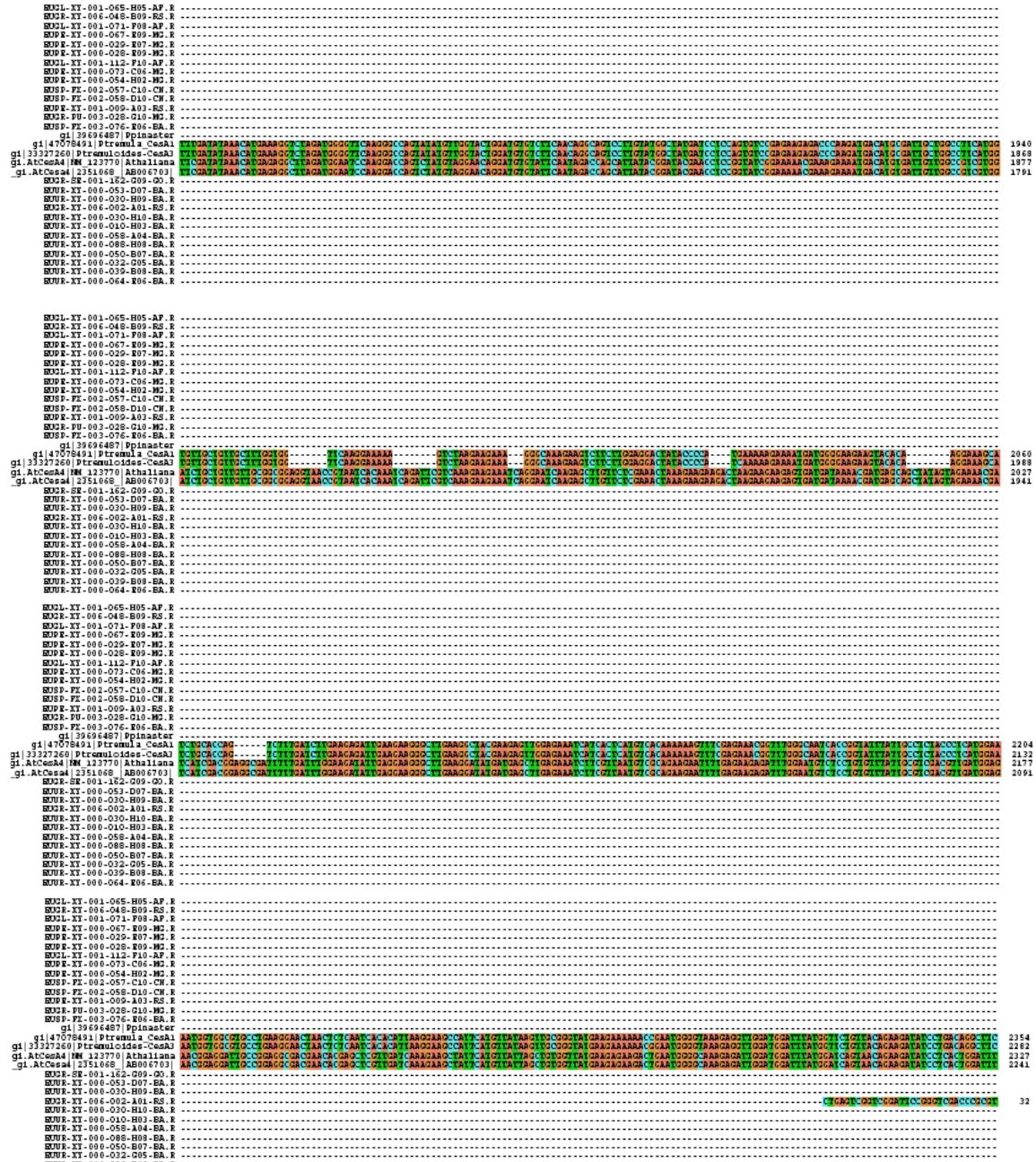
ANEXO 1

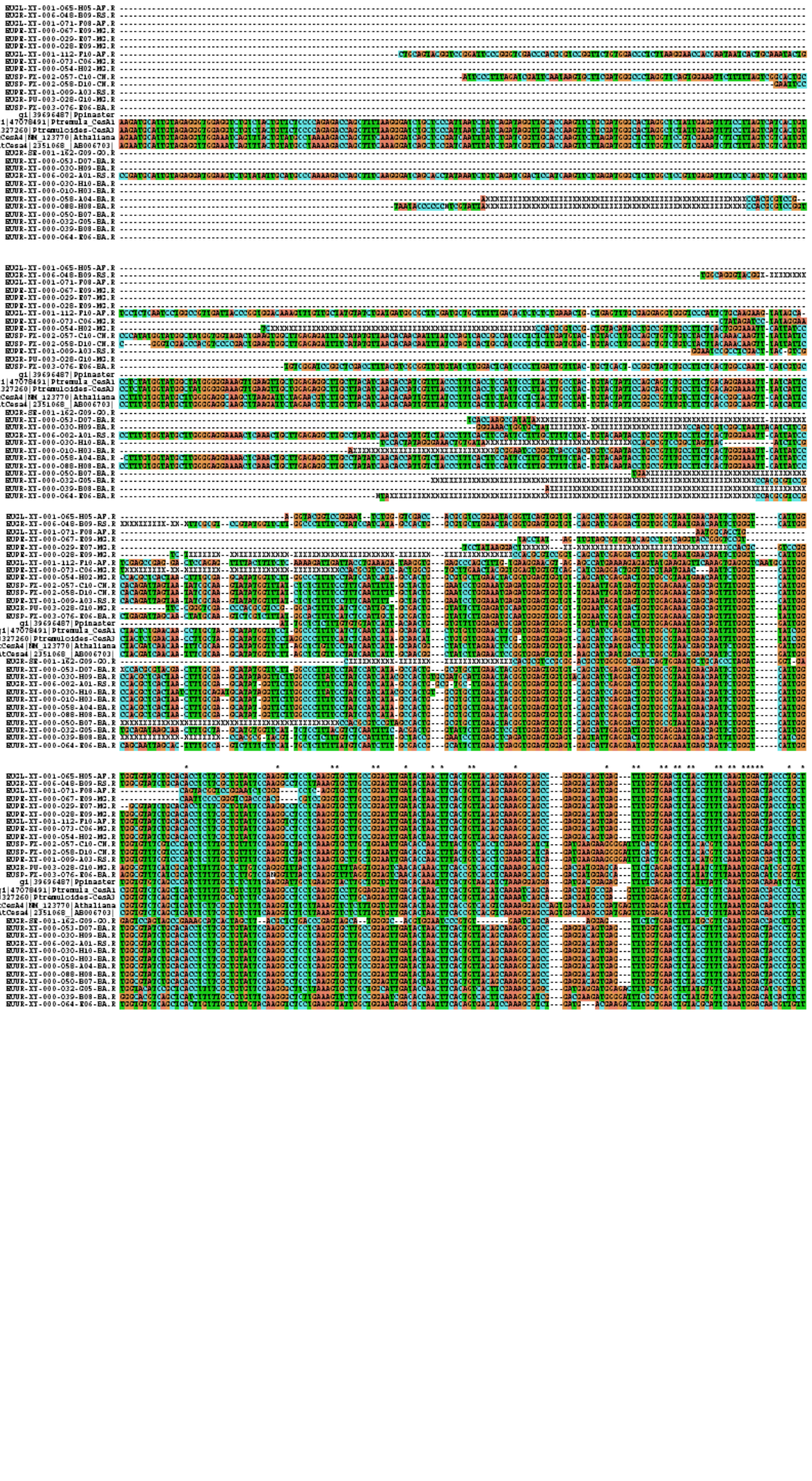
Alinhamento de ESTs de *Eucalyptus* e seqüências de *Pinus sp.* e *Populus sp.* ortólogos à *AtCesA4* de *A. thaliana.*

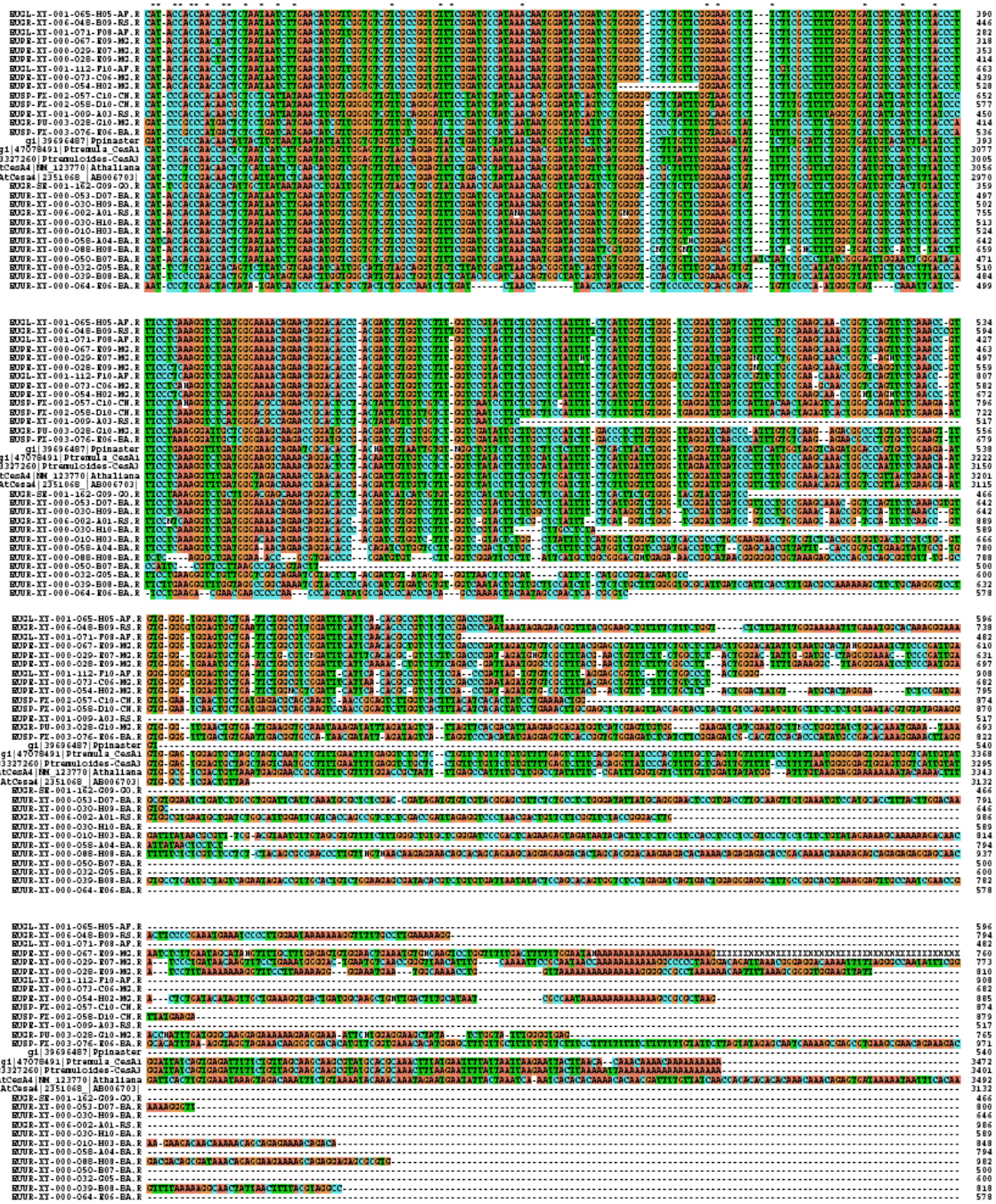






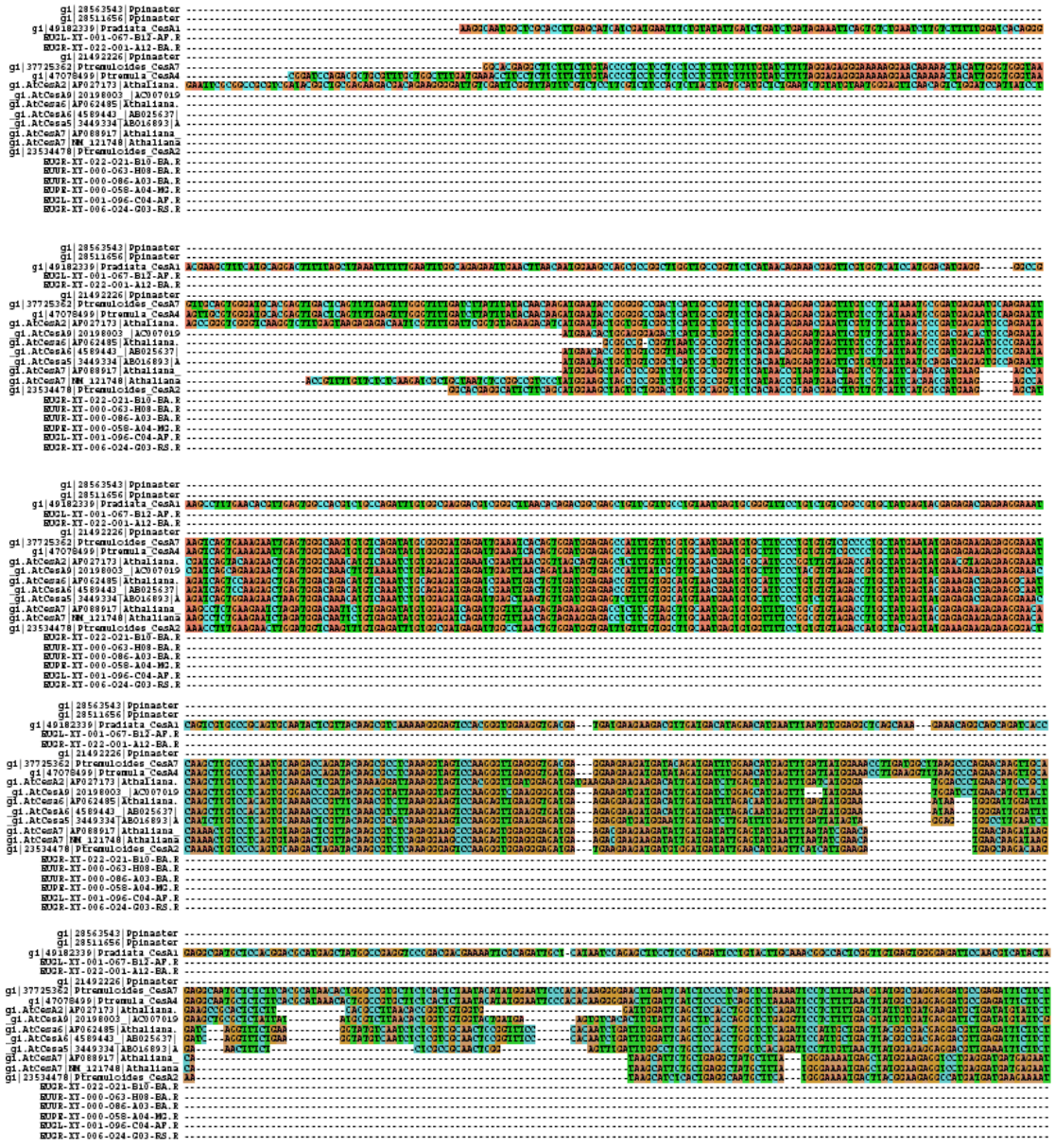


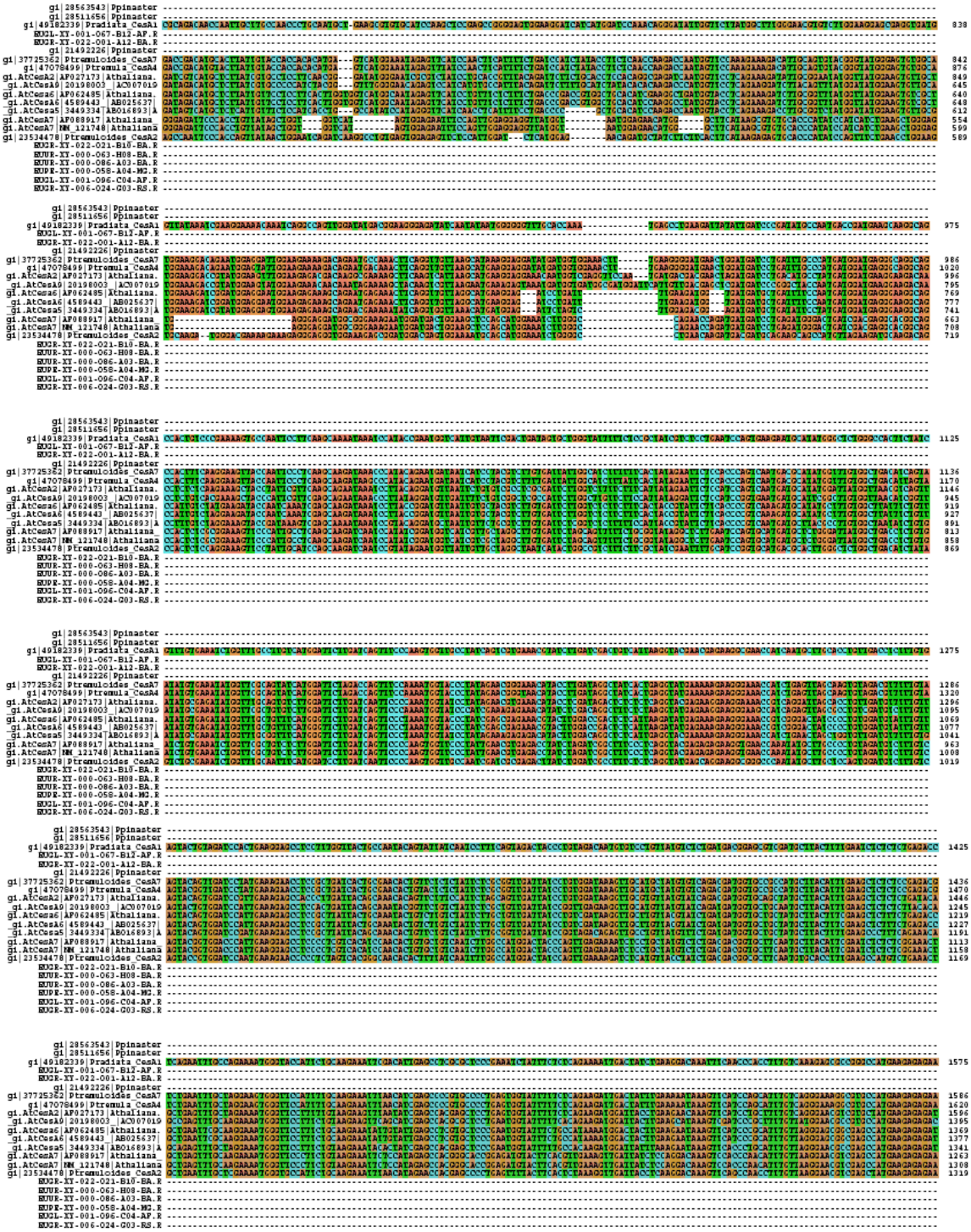


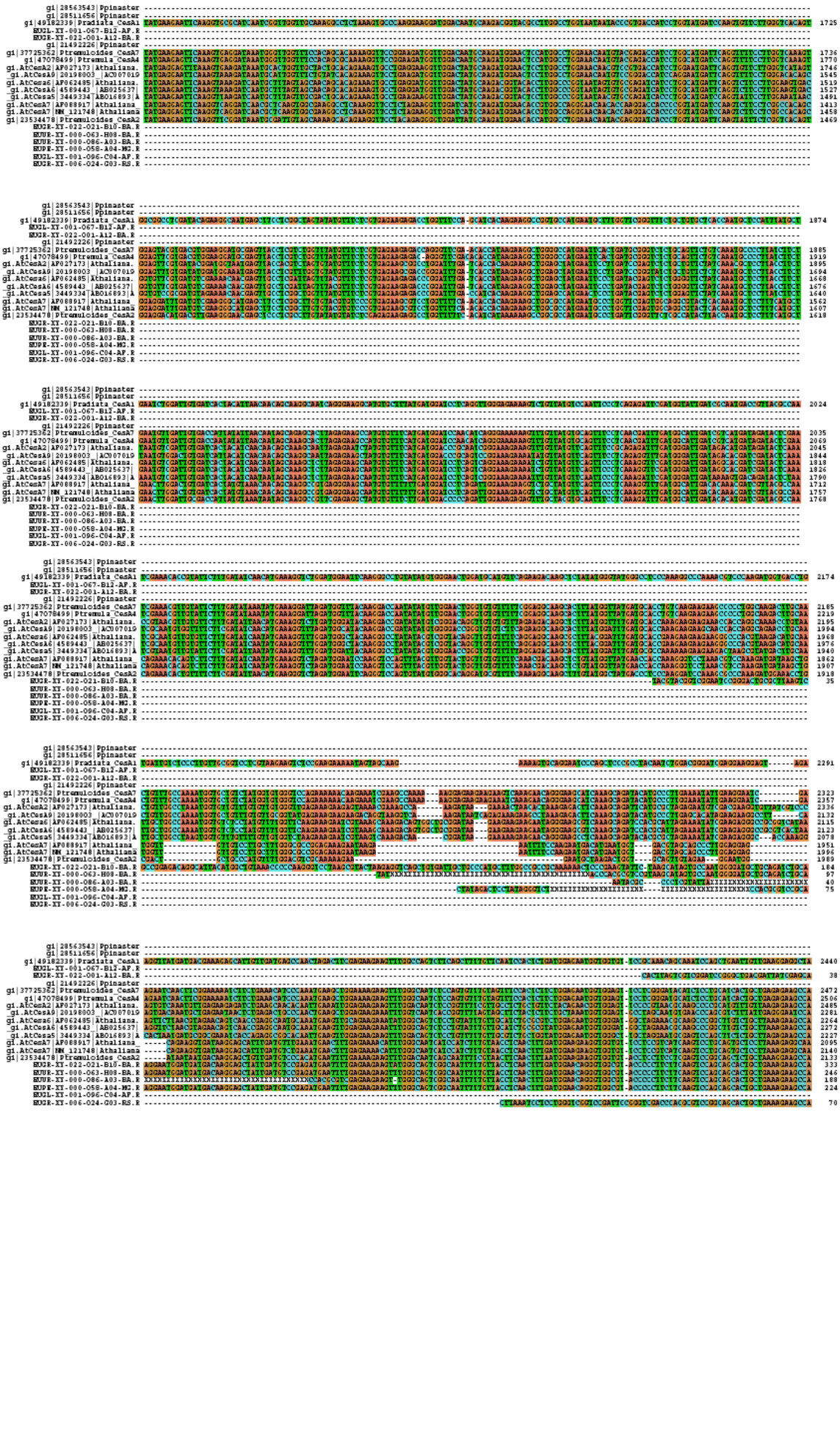


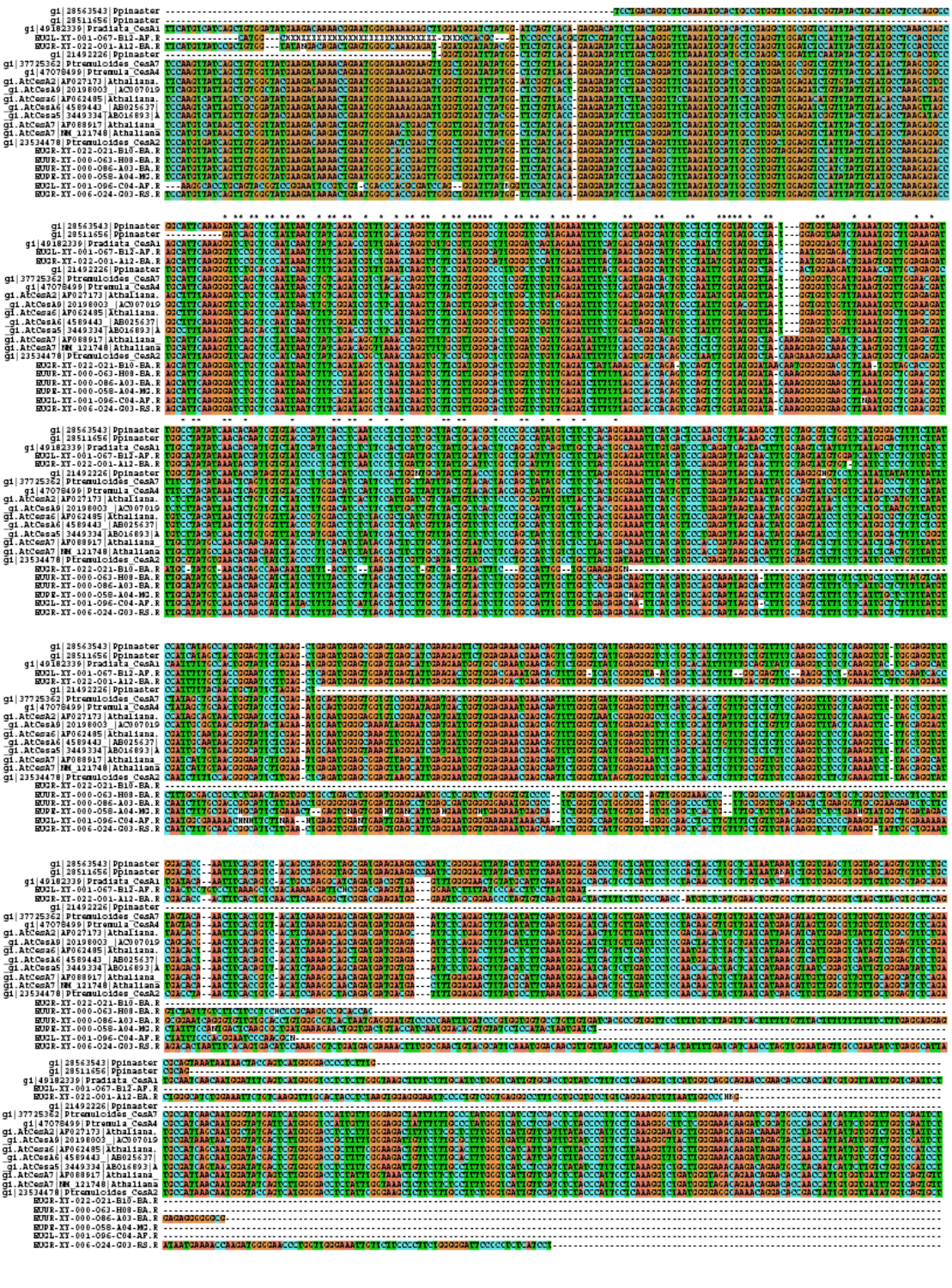
ANEXO 2

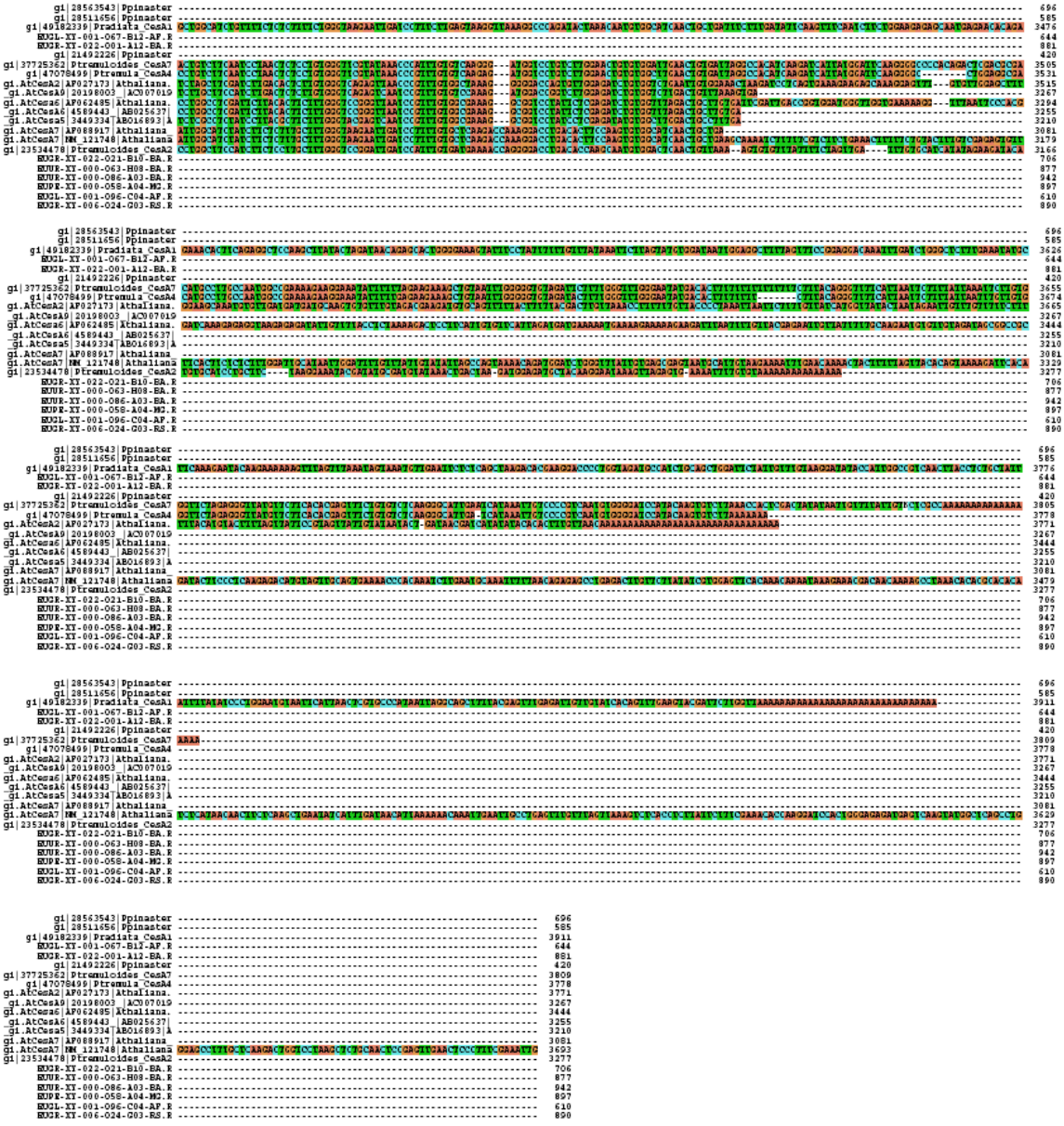
Alinhamento de ESTs de *Eucalyptus* e seqüências de *Pinus sp.* e *Populus sp.* ortólogos à *AtCesA7* de *A. thaliana.*





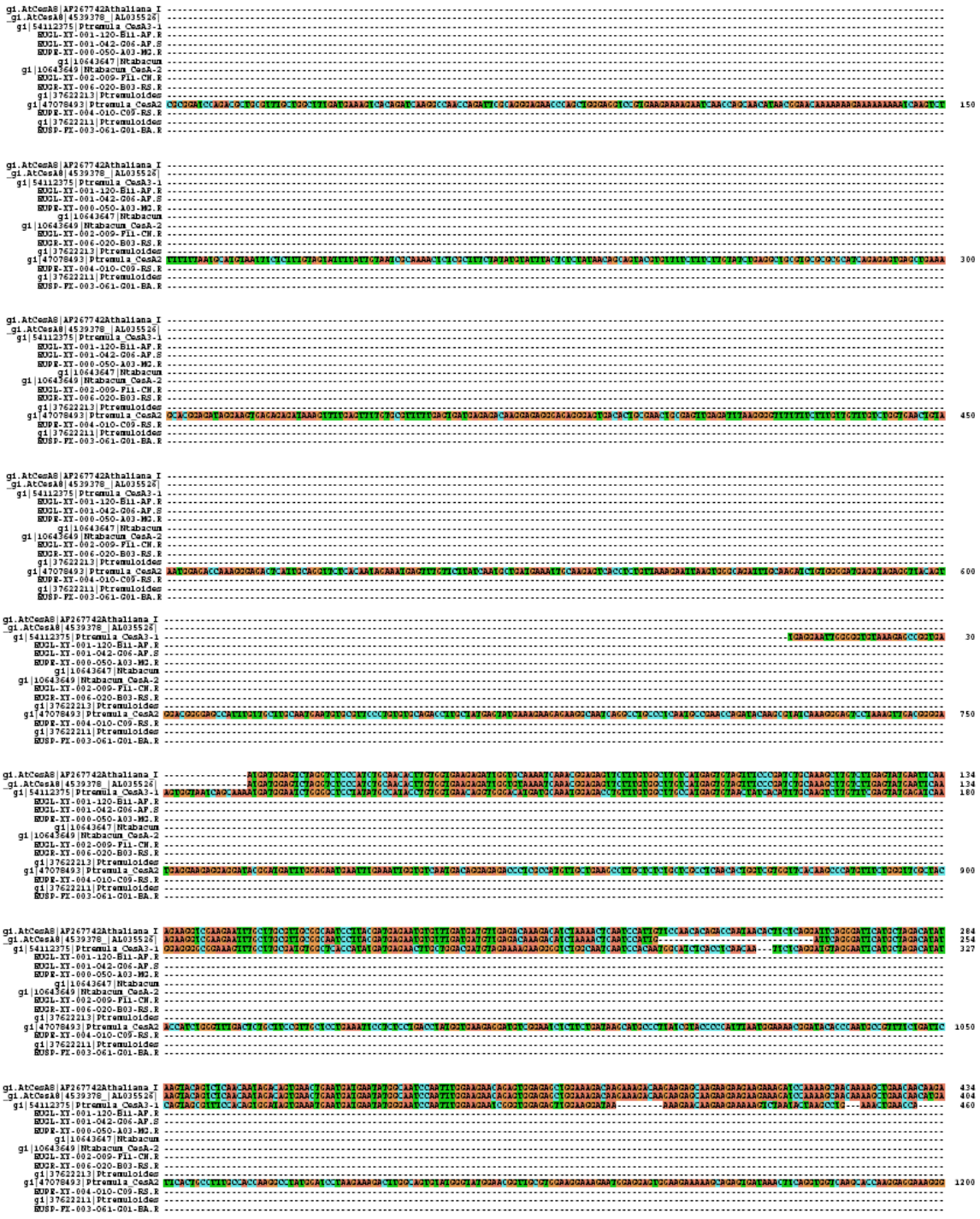


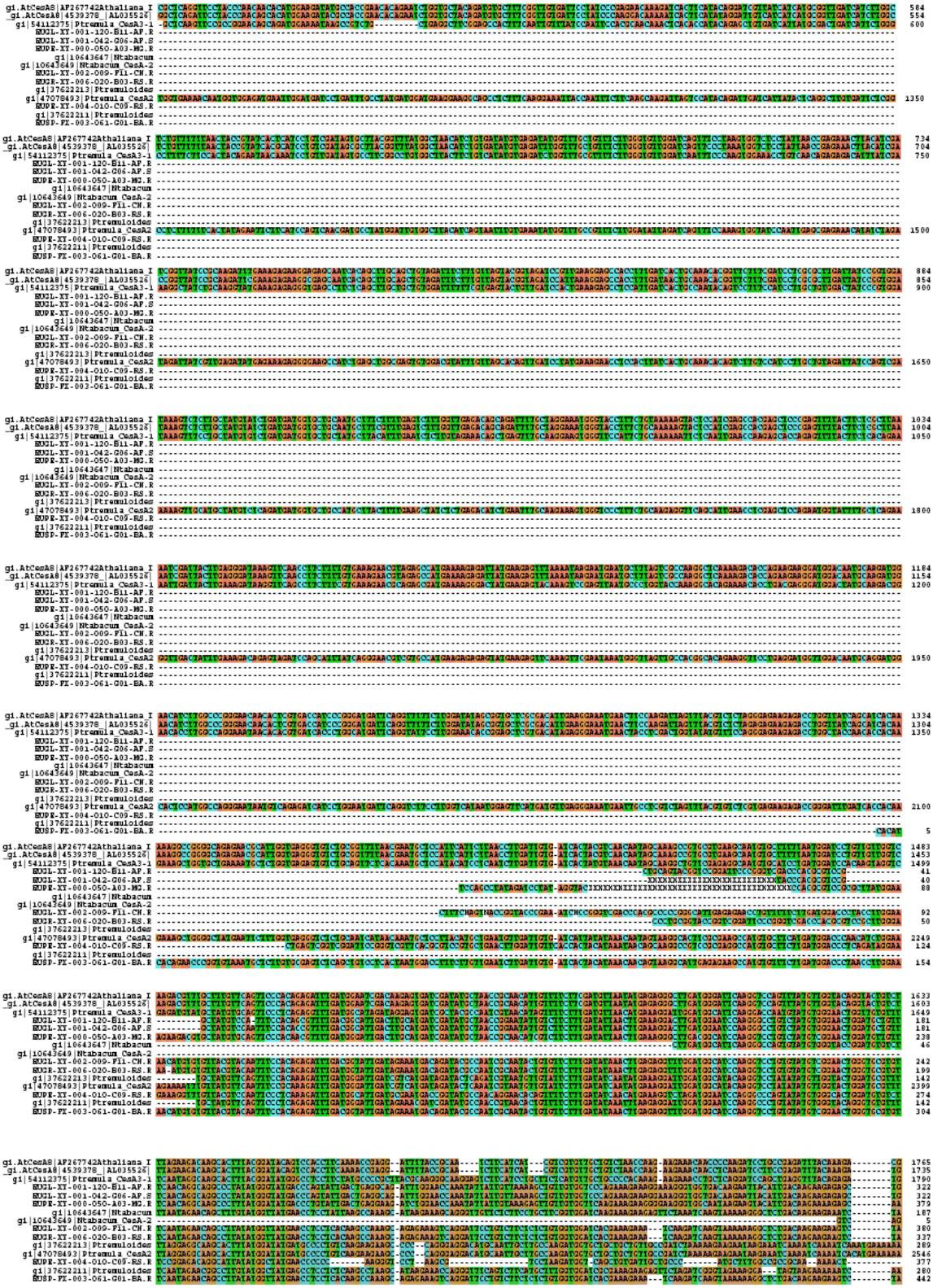


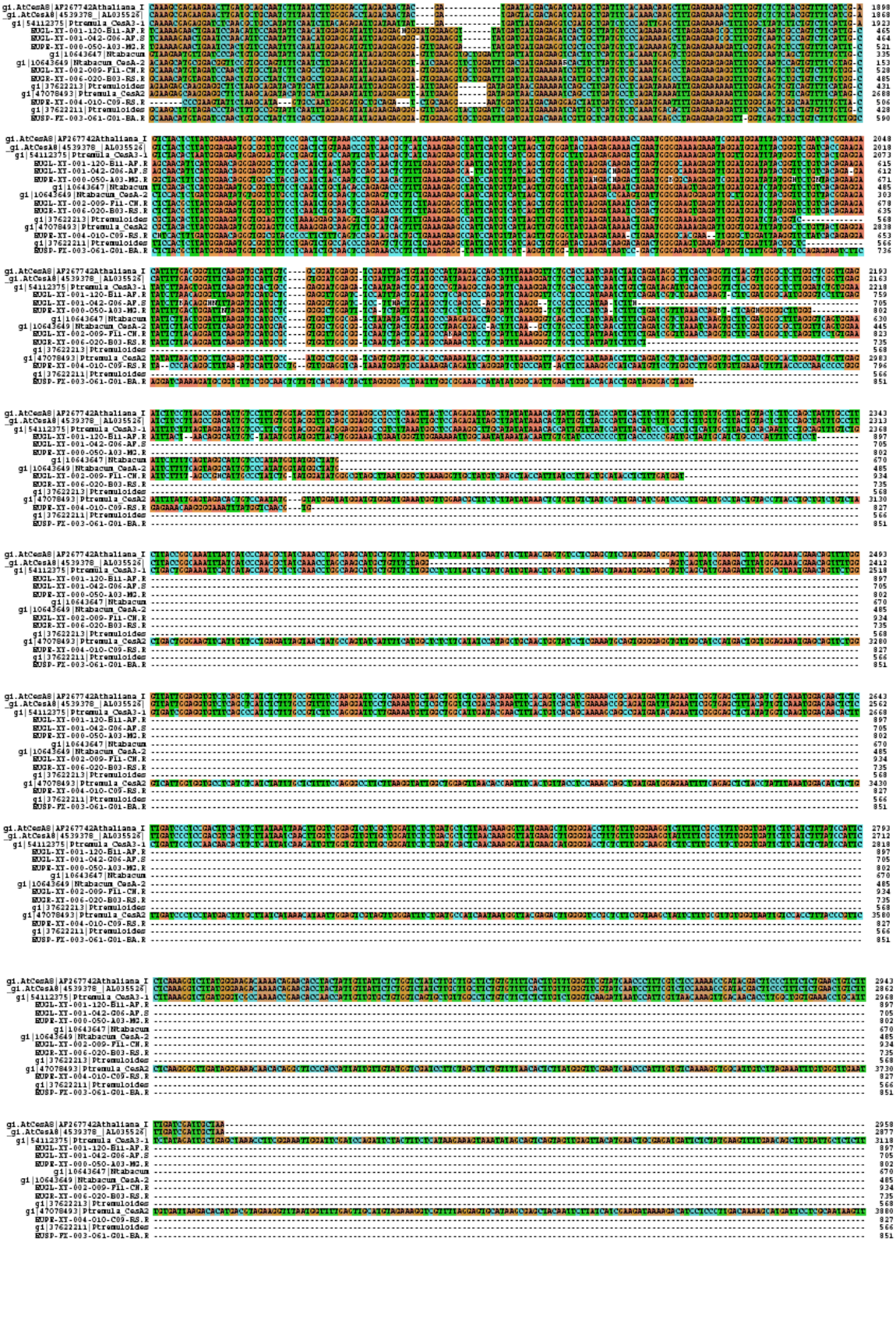


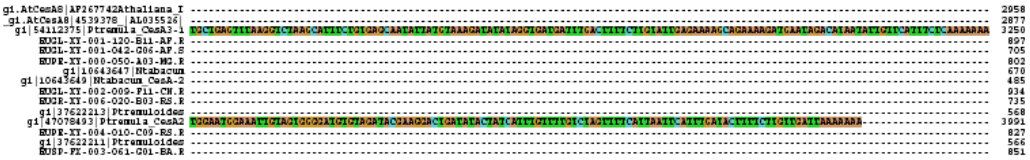
ANEXO 3

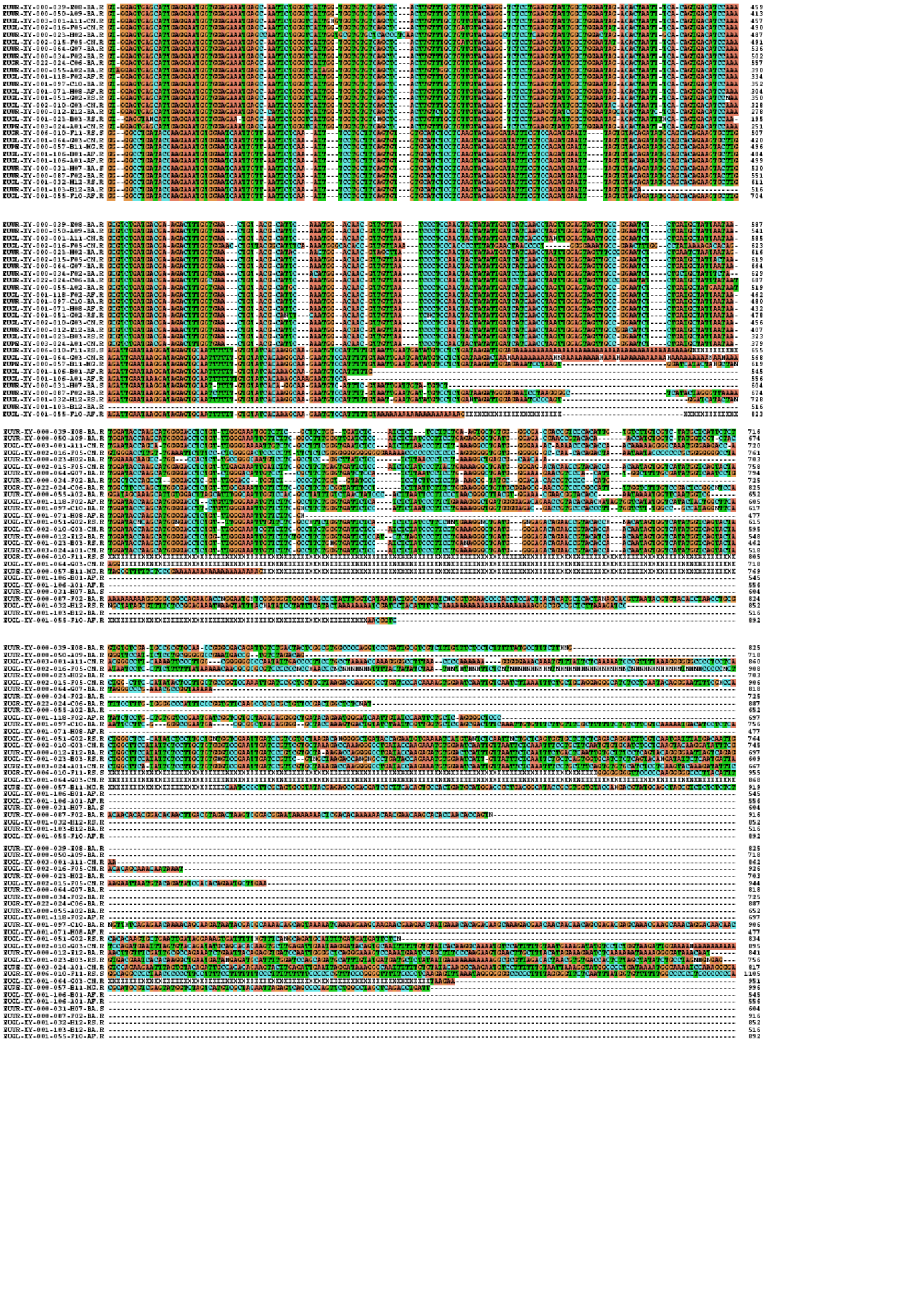
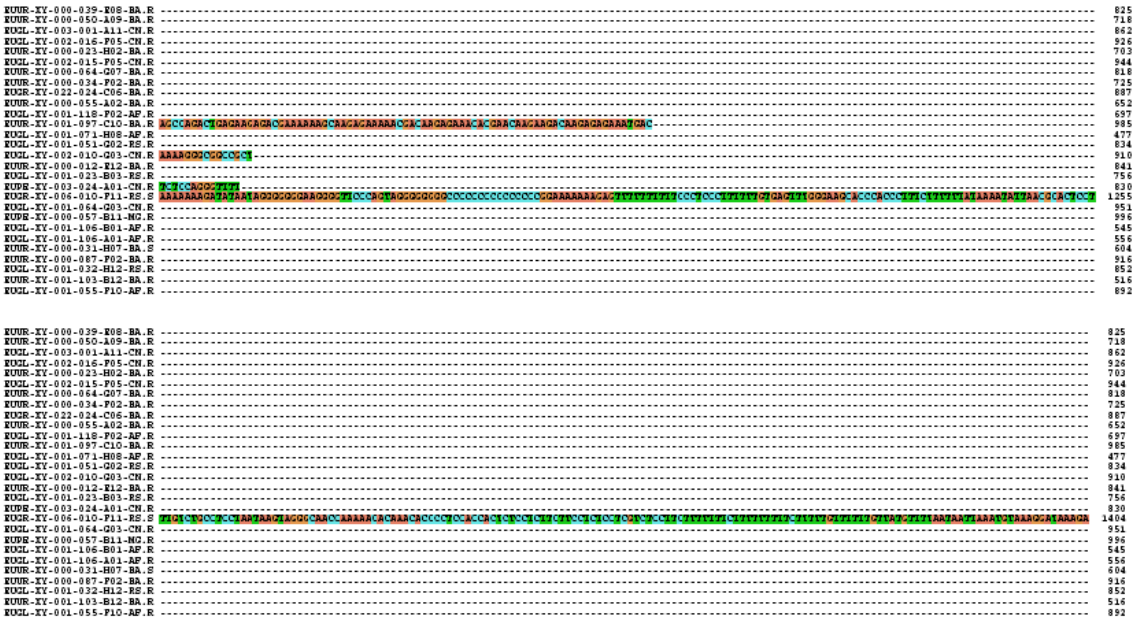
Alinhamento de ESTs de *Eucalyptus* e seqüências de *Pinus sp.* e *Populus sp.* ortólogos à *AtCesA8* de *A. thaliana.*








gi.AtCesA8|AF267742Athaliana_I 2958
gi.AtCesA8|4539378_|AL035526| 2877
gi|54112375|Ptremula_CesA3-1 3250
EU2L-XY-001-120-E11-AF.R 897
EU2L-XY-001-042-C06-AF.S 705
EUPR-XY-000-050-A03-M2.F 802
gi|10643647|Ntabacum 670
gi|10643649|Ntabacum_CesA-2 485
EU2L-XY-002-009-F11-CM.R 934
EU2R-XY-006-020-B03-RS.R 735
gi|37622213|Ptremuloides 568
gi|47078493|Ptremula_CesA2 3991
EUPR-XY-004-010-C09-RS.R 827
gi|37622211|Ptremuloides 566
EUSP-FZ-003-061-C01-BA.R 851

ANEXO 4

Alinhamento entre ESTs de *Eucalyptus* homólogos à genes *AtCesA* de *A. thaliana* responsáveis pela produção de celulose na parede secundária das plantas.